Alibaba Cloud
Apsara Stack Enterprise

User Guide - Cloud Essentials and Security

Version: 1909, Internal: V3.8.1

Issue: 20200116

Legal disclaimer

Alibaba Cloud reminds you to carefully read and fully understand the terms and conditions of this legal disclaimer before you read or use this document. If you have read or used this document, it shall be deemed as your total acceptance of this legal disclaimer.

1. **You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.**
2. **No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company, or individual in any form or by any means without the prior written consent of Alibaba Cloud.**
3. **The content of this document may be changed due to product version upgrades, adjustments, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and the updated versions of this document will be occasionally released through Alibaba Cloud-authorized channels. You shall pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.**
4. **This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides the document in the context that Alibaba Cloud products and services are provided on an "as is", "with all faults" and "as available" basis. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not bear any liability for any errors or financial losses incurred by any organizations, companies, or individuals arising from their download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, bear responsibility for any indirect, consequent**

ial, exemplary, incidental, special, or punitive damages, including lost profits arising from the use or trust in this document, even if Alibaba Cloud has been notified of the possibility of such a loss.

5. By law, all the contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of this document shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates.
6. Please contact Alibaba Cloud directly if you discover any errors in this document.

Document conventions

Style	Description	Example
	A danger notice indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	 **Danger:** Resetting will result in the loss of user configuration data.
	A warning notice indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	 **Warning:** Restarting will cause business interruption. About 10 minutes are required to restart an instance.
	A caution notice indicates warning information, supplementary instructions, and other content that the user must understand.	 **Notice:** If the weight is set to 0, the server no longer receives new requests.
	A note indicates supplemental instructions, best practices, tips, and other content.	 **Note:** You can use Ctrl + A to select all files.
>	Closing angle brackets are used to indicate a multi-level menu cascade.	Click Settings > Network > Set network type.
Bold	Bold formatting is used for buttons, menus, page names, and other UI elements.	Click **OK**.
`Courier font`	Courier font is used for commands.	Run the `cd /d C:/window` command to enter the Windows system folder.
Italic	Italic formatting is used for parameters and variables.	`bae log list --instanceid` *`Instance_ID`*
[] or [a|b]	This format is used for an optional value, where only one item can be selected.	`ipconfig [-all|-t]`

Style	Description	Example
{} or {a|b}	**This format is used for a required value, where only one item can be selected.**	`switch {active|stand}`

Contents

1 Apsara Stack console

1.1 What is the Apsara Stack console?

The Apsara Stack console is a service capability platform that is based on the Alibaba Cloud Apsara Stack platform and is customized for government and enterprise customers. This platform focuses on improving IT management and solving operation problems for you, and is dedicated to providing a leading service capability platform of the cloud computing industry. It provides large-scale and cost-effective one-stop cloud computing and big data services for customers in many industries, such as government, education, healthcare, finance, and enterprise.

Overview

The Apsara Stack console builds an Apsara Stack platform that supports different business modes for governments and enterprises, which simplifies the management and deployment of physical and virtual resources, helps you easily and rapidly establish your own business system with higher resource utilization and lower Operation and Maintenance (O&M) costs. It shifts your attention from operation and O&M to business, brings the Internet economic mode to government and enterprise customers, and builds a brand new ecological chain that is based on cloud computing.

1.2 Log on to the Apsara Stack console

This topic describes how to log on to the Apsara Stack console as cloud product users.

Prerequisites

- **Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://***IP address or domain name of the Apsara Stack console***/manage.**
- **We recommend that you use the Chrome browser.**

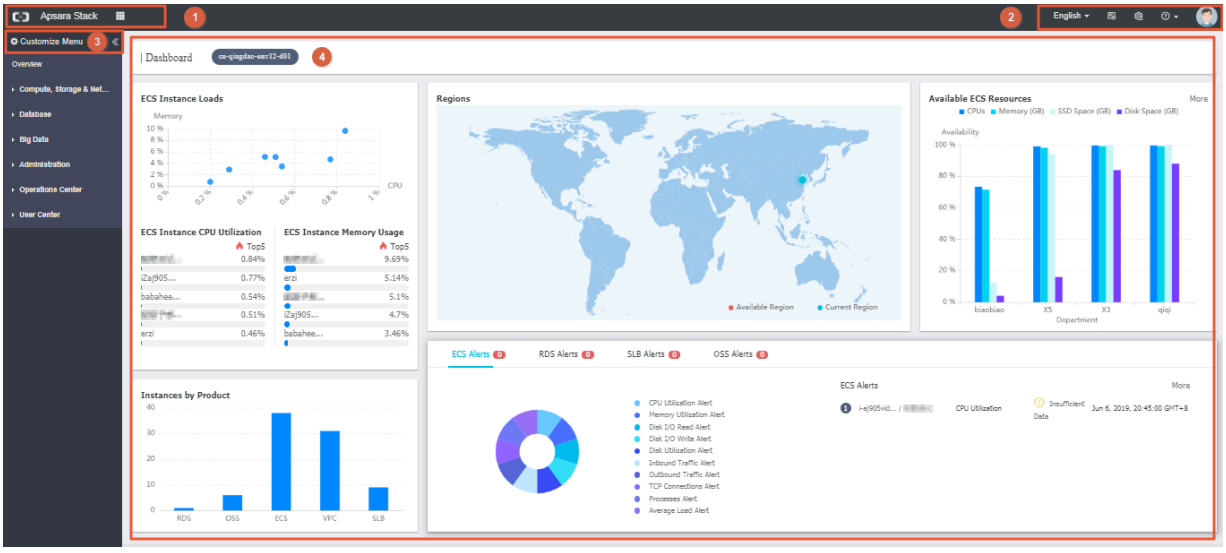
Procedure

1. **Open your browser.**
2. **In the address bar, enter the access address of the Apsara Stack console in the format of http://`IP address or domain name of the Apsara Stack console`/manage, and then press Enter.**
3. **Enter the correct username and password.**
   - **The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.**
   - **You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).**
4. **Click LOGIN to go to the Dashboard page.**

1.3 Web page introduction

The Web page of the Apsara Stack console consists of three areas: main menu bar, information area of the current logon user, and operation area.

Figure 1-1: Apsara Stack console



For more information about the functional areas of the Web page, see *Table 1-1: Functional areas of the Web page*.

Table 1-1: Functional areas of the Web page

Area		Description
1	Main menu bar	· Apsara Stack: Click Apsara Stack to go back to the dashboard page when you are in another page. · Capability Center: Click  to choose products or services as required. It contains the following modules: - Compute, Storage & Networking: manages all types of basic cloud products and resources. - Database: manages all types of database products and resources. - Big Data: manages all types of big data products and resources. - Administration: manages the CloudMonitor, System Reports, Operation Log, and Task Center of the system. - Operations Center: manages resource allocation of the system. - User Center: manages the departments, projects, roles, users, and logon policies of the system. **Note:** The menu bar varies with different roles. See your menu bar for relevant functions.

Area		Description
2	Information area of the current logon user	: Click this to select language. · : Click this to display the history and most frequently accessed menu items. · : Click this to go to the System Configuration page. · : Place your pointer over this icon. Select User Guide to go to the Content Center page. · Click your avatar and select items to go to Personal Information page, Theme page, and Log Off page. On the Personal Information page, you can: - View your basic information. - Modify your information. - Modify your profile picture. - Modify your logon password. - View the AccessKey. - View the third-party AccessKey.
3	Customize Menu	Configures your common menu items. : Click this to configure the displayed items in the left-side navigation pane. : Click this to expand and collapse the left-side navigation pane.
4	Operation area	Dashboard: displays the overview and monitoring status of resources in each department in the Apsara Stack console.

1.4 Configuration of system initialization

1.4.1 Configuration instruction

The administrator must complete a series of basic configurations according to the configuration process, such as creating departments, roles, projects, users, and initializing resources before using the Apsara Stack console.

The Apsara Stack console follows service principles to perform unified management for users, roles, departments, and projects related to cloud data center, which allows you to grant different resource access permissions to different users.

- **Department**

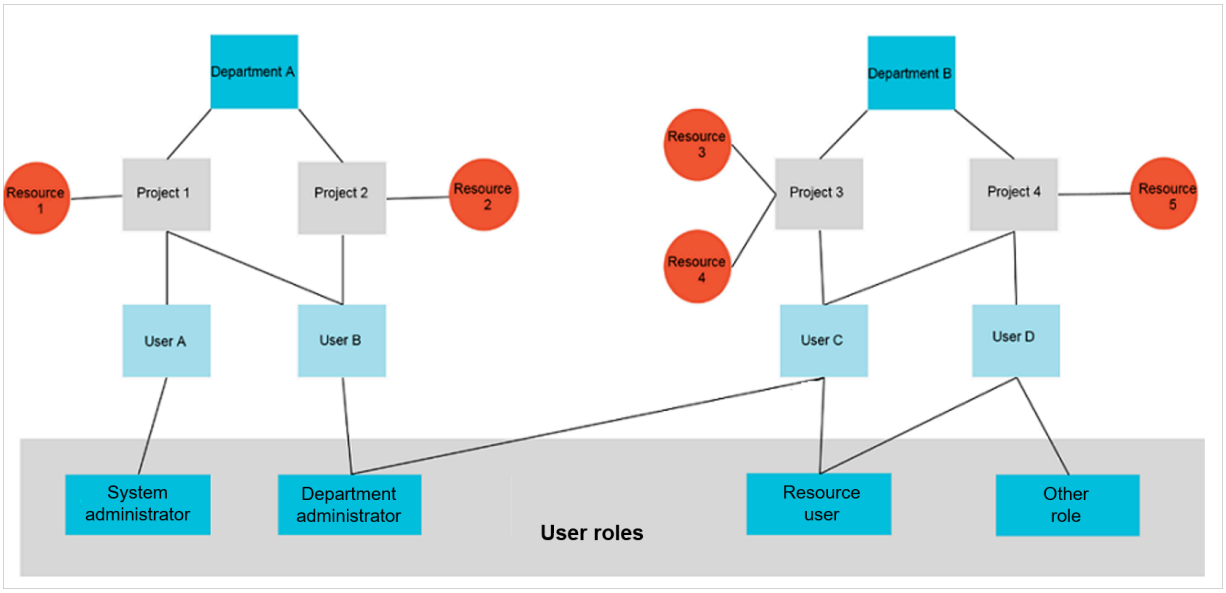
  After the Apsara Stack console is deployed, the system creates a root department by default. You can create departments under the root department.

  The system displays departments hierarchically and you can create sub-departments under each department.
- **Role**

  A collection of access permissions. When creating users, you must assign different roles to users to meet their access control requirements on the system.
- **Project**

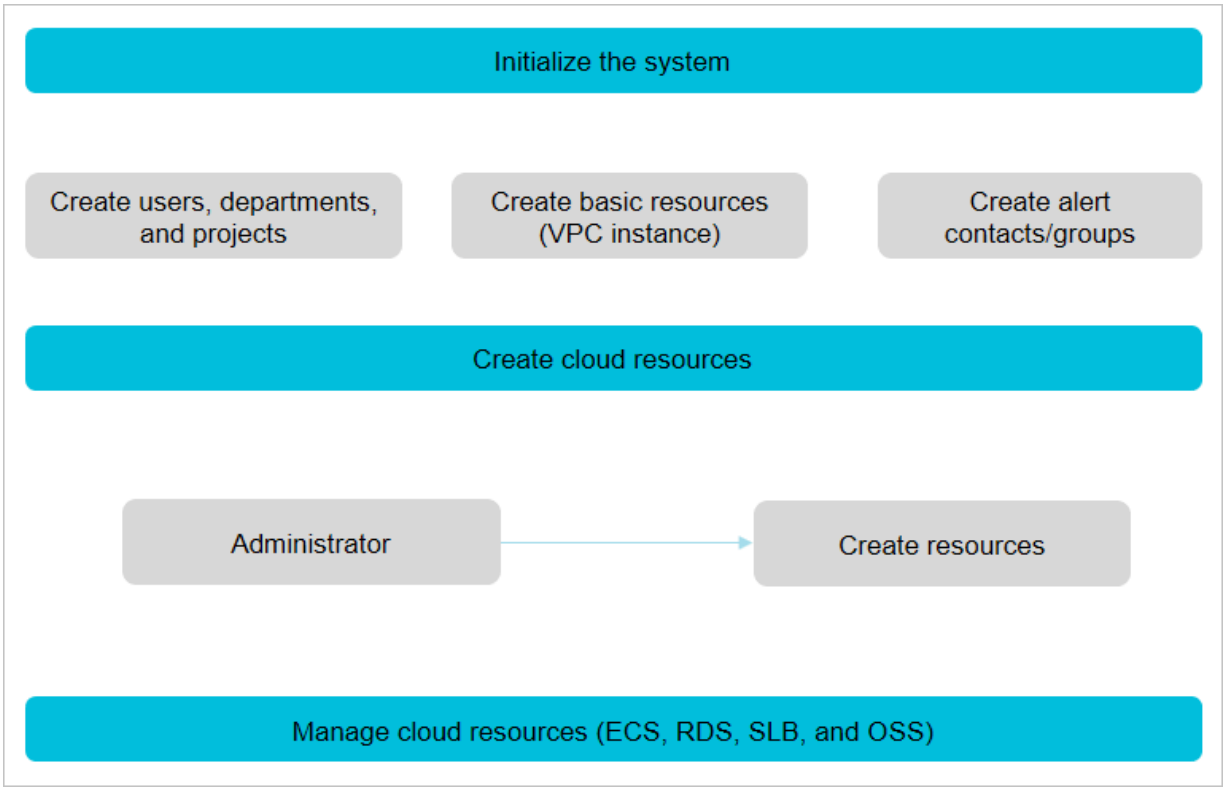
  A container where resources are stored. All resources must be created under a corresponding project.

The relationships among departments, users, projects, roles, and cloud resources are as follows.



After initializing the system, the administrator initializes resources, creating cloud resources for project members to use.

Configuration process



The main operations available in the Apsara Stack console are as follows:

- **Initialize the system: completes the basic system configurations such as creating departments, projects, users, basic resources (Virtual Private Cloud (VPC) instances), alert contacts, contact groups, and pricing the resources.**

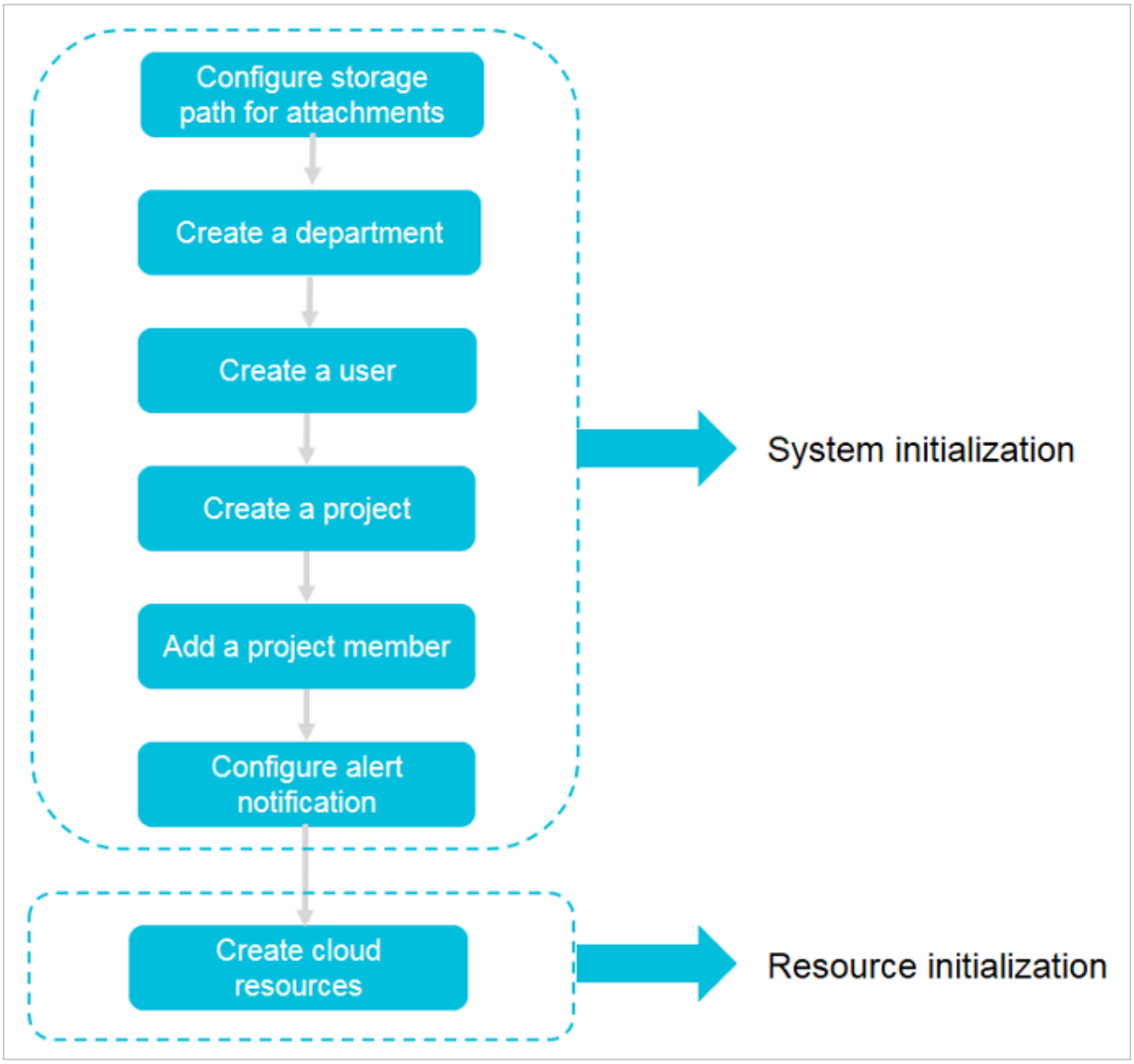
- **Create cloud resources: The administrator directly creates resources as required.**
- **Manage cloud resources: manages resources, such as starting, using, applying , and releasing resources, and changing resource configurations and applying quotas.**

1.4.2 Configuration process

This topic describes the initial configuration process of the system.

Before using the Apsara Stack console, the administrator must complete the initial configuration of the system. The process is as shown in the following figure.

Figure 1-2: Initial configuration process of the system



1. *Configure the storage path for attachments*

   Configure the storage path for uploaded attachments.

2. *Create a department*

   Create a department to store projects and the resources in the projects.

3. *Create a user*

   The administrator can create users and assign different roles to users to meet their access control requirements on the system.

4. *Create a project*

   Create a project before you apply for resources.

5. *Add a project member*

   Add a user to a project.

6. *Configuration of alert notification*

   Configure the alert notification to allow alert contacts to receive alert notificati ons by email, SMS, and DingTalk when alerts are triggered in real time.

7. **Create cloud resources**

   The administrator can create cloud product instances in the console of each cloud product according to the project requirements. For more information about how to create cloud product instances, see the detailed introduction of each cloud product.

1.5 Resource management

1.5.1 Quota management

1.5.1.1 Quota parameters

This topic describes the quota parameters of products.

The administrator can configure resource quotas for departments. The department administrator can create resources directly within quotas. When the quotas are used up, the system forbids the department administrator to create more resources. If you want to create more, you must add resource quotas for the current department.

You can create limitless resources if you do not configure quotas.

OSS

Parameter	Description
Total OSS Instances	The total number of buckets that you can configure for Object Storage Service (OSS).
Total OSS Capacity (GB)	The total size of buckets that you can configure for OSS.

ECS

Parameter	Description
Total CPU Quota (Cores)	The total number of CPU cores that you can configure for Elastic Compute Service (ECS) and the number of used cores.
Total Memory Quota (GB)	The total memory size that you can configure for ECS.
Total SSD Quota (GB)	The total SSD size that you can configure for ECS.
Total Disk Quota (GB)	The total number of cloud disks that you can configure for an ECS instance.

EGS

Parameter	Description
Total CPU Quota (Cores)	The total number of CPU cores that you can configure for Elastic GPU Service (EGS).
Total Memory Quota (GB)	The total memory size that you can configure for EGS.
Total GPU Quota (GB)	The total size of GPU cores that you can configure for EGS.

RDS (including primary instances and read-only instances)

Parameter	Description
Total CPU Quota (Cores)	The total number of CPU cores that you can configure for ApsaraDB for RDS (MySQL/PPAS/PostgreSQL) and the number of used cores.

Parameter	Description
Total Memory Quota (GB)	The total memory size that you can configure for ApsaraDB for RDS (MySQL/PPAS/PostgreSQL).
Total Storage Quota (GB)	The total storage size that you can configure for ApsaraDB for RDS (MySQL/PPAS/PostgreSQL).

SLB

Parameter	Description
Total External IP Addresses	The total number of external IP addresses that you can configure for Server Load Balancer (SLB).
Total Internal IP Addresses	The total number of internal IP addresses that you can configure for SLB.

AnalyticDB for PostgreSQL

Parameter	Description
Total CPU Quota (Cores)	The total number of CPU cores that you can configure for AnalyticDB for PostgreSQL and the number of used cores.
Total Memory Quota (GB)	The total memory size that you can configure for AnalyticDB for PostgreSQL.
Total Storage Quota (GB)	The total storage size that you can configure for AnalyticDB for PostgreSQL.

MongoDB

Parameter	Description
Total CPU Quota (Cores)	The total number of CPU cores that you can configure for ApsaraDB for MongoDB and the number of used cores.
Total Memory Quota (GB)	The total memory size that you can configure for ApsaraDB for MongoDB.
Total Storage Quota (GB)	The total storage size that you can configure for ApsaraDB for MongoDB.

VPC

Parameter	Description
VPCs	The total number of Virtual Private Cloud (VPC) instances that you can configure for VPC.

EMR

Parameter	Description
Total CPU Quota (Cores)	The total number of CPU cores that you can configure for E-MapReduce (EMR).

MaxCompute

Parameter	Description
CU Quota	The total number of CUs that you can configure for MaxCompute.
Total Storage Quota (GB)	The total storage size that you can configure for MaxCompute.

QuickBI

Parameter	Description
Users Quota	The total number of users that you can configure for QuickBI.

NAS

Parameter	Description
Storage Quota	The total space size that you can configure for Network Attached Storage (NAS).

Redis

Parameter	Description
Total Memory Quota (GB)	The total memory size that you can configure for KVStore for Redis.

SLS

Parameter	Description
Storage Quota	The total space size that you can configure for Log Service.

DataWorks

Parameter	Description
Total Table Quota (Tables)	The total number of tables that you can configure for DataWorks. Note: The table quota must be an integer multiple of 5,000.
Total Storage Space (TB)	The total storage size that you can configure for DataWorks. Note: This parameter is automatically configured after you configure the table quota. You are not required to configure it.

OTS

Parameter	Description
Total Storage Quota (GB)	The total storage size that you can configure for Table Store (OTS).

Stream Compute

Parameter	Description
Total CU	The total number of CUs that you can configure for Realtime Compute.

1.5.1.2 View total and used quotas

The administrator can view the total quotas, used quotas, and used resources of cloud resources in different departments and regions.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click  , then choose Operations Center > Quota Management.
3. Click the Overview tab.
4. Select the region, department, and product. Then, the system displays the overview of quotas.

   In the search result, view the total quotas, used quotas, and used resources of a certain product.

Parameter	Description
Total quotas	The total quotas of the selected department.
Used quotas	The combination of used resources of the selected department and quotas allocated to its sub-departments.
Used resources	The used resources of the selected department.

1.5.1.3 Create cloud resource quotas

The Apsara Stack console supports configuring quotas to allocate resources reasonably among departments.

Context

You can create quotas for a sub-department. If the parent department (except a level-1 parent department) has a quota, the result that the quota of the parent department minus the quotas of other sub-departments is the maximum quota that you can configure for the sub-department. The result cannot be smaller than the amount of resources already created for the sub-department.

This topic describes how to create quotas for Elastic Compute Service (ECS). You can create quotas for other cloud resources in a similar way.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.

2. In the top navigation bar, click  , then choose Operations Center > Quota Management.
3. Click the Quota tab.
4. Select the Region and Department.
5. Locate the section of ECS, the product for which you are about to create quotas.
6. In the upper-right corner, click the  icon.
7. Configure the total quotas and then click the  icon.

   For more information about the quota parameters, see *Quota parameters*.

1.5.1.4 Modify quotas

The administrator can modify cloud resource quotas based on the department requirements.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click  , then choose Operations Center > Quota Management.
3. Click the Quota tab.
4. Select the Region and Department.
5. Select the product whose quotas you want to modify.

   The system displays the product quota section.
6. In the upper-right corner of the product quota section, click the  icon.
7. Reenter the quotas and then click the  icon.

1.5.1.5 Delete quotas

The administrator can delete quotas as required.

Prerequisites

Before deleting quotas, make sure that all sub-departments of the selected department do not have any quotas.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click  , then choose Operations Center > Quota Management.
3. Click the Quota tab.
4. Select the Region and Department.
5. Select the product whose quotas you want to delete.

   The system displays the product quota section.
6. In the upper-right corner of the product quota section, click the  icon.

   Values in the product quota section are cleared.

1.5.2 View and export the resources overview of projects

The cloud resource overview displays the numbers of resources and alerts in each department and project in the Apsara Stack console in the format of lists.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or a user.
2. In the left-side navigation pane, click Overview to go to the Products Overview page.
3. Click the Resource Overview tab.

   On the Resource Overview tab, you can quickly view the numbers of resources and alerts in each department and project to learn about the distribution and running conditions of resources.

   > **Note:**
   > Click a number on the page and then you are redirected to the corresponding resource page where you can view the detailed resource information.
4. Click Export to go to the Resource Reports tab of the system reports.

1.5.3 Configuration of resource notifications

1.5.3.1 Configure resource notification objects

You can configure the resource notification objects to receive emails and SMS notifications from the Apsara Stack console when resources are created or deleted.

Context

You can add users as notification objects by configuring the resource notifications . If resources are created or deleted in the Apsara Stack console, the system sends emails and SMS notifications to the configured notification objects. Users who are added as notification objects can keep up with the resource usage.

Note:

The Apsara Stack console can send resource notifications for Elastic Compute Service (ECS), Object Storage Service (OSS), ApsaraDB for RDS, and Server Load Balancer (SLB) resources.

Procedure

1. *Log on to the Apsara Stack console* **as a system administrator.**
2. **In the upper-right conner, click the**  **icon to go to the System Configuration page.**
3. **Click the Resource Notification Configuration tab.**
4. **Click Add in the upper-right corner of the table.**
5. **In the displayed Add User dialog box, select users in the Available Users field and click the icon to add them to Selected Users.**
6. **Click Add to complete the configurations.**

1.5.3.2 View resource notification objects

You can view information about resource notification objects.

Procedure

1. *Log on to the Apsara Stack console* **as a system administrator.**
2. **In the upper-right conner, click the**  **icon to go to the System Configuration page.**
3. **Click the Resource Notification Configuration tab.**

4. On the Resource Notification Configuration tab, view the resource notification objects in the Apsara Stack console.
5. Optional: You can click the username of a resource notification object to view the detailed information of this notification object.

1.5.3.3 Delete a resource notification object

The administrator can remove users who are no longer required to be notified of new changes from resource notification objects because of business changes or other reasons.

Procedure

1. *Log on to the Apsara Stack console* as a system administrator.
2. In the upper-right corner, click the  icon to go to the System Configuration page.
3. Click the Resource Notification Configuration tab.
4. Find the user to be deleted. Click the  icon in the Actions column and select Delete.
5. In the displayed dialog box, click OK.

1.6 Alert management

1.6.1 Overview

CloudMonitor provides real-time monitoring, alert, and notification services of resources to protect your products and business.

Currently, CloudMonitor can monitor metrics of Elastic Compute Service (ECS), Server Load Balancer (SLB), ApsaraDB for RDS, Object Storage Service (OSS), and KVStore for Redis.

You can use the monitoring metrics of cloud products to set alert rules and notification policies to keep up with the running conditions and performance of instance resources for product services. Consider a scale-up in time after receiving an insufficient resource alert.

CloudMonitor has the following functions:

- **Automatic monitoring: The monitoring is automatically started based on your created ECS resources or Auto Scaling groups. You are not required to start it manually or install any plug-ins. You can view the monitoring data of specific instances on the monitoring page after applying for resources.**
- **Flexible alert: You can configure alerts flexibly, such as setting alerts and thresholds for monitoring metrics, pausing and starting alerts.**
- **Real-time notification: You can set the alert notification to receive notifications by SMS or email in real time. If the status of an alert rule changes, such as alerts are triggered, data is insufficient, or alerts are cleared, the system informs you by sending SMS or email.**

1.6.2 Alert contacts

1.6.2.1 Create an alert contact

You can create an alert contact to receive alert notifications.

Context

An alert contact is a person who receives alert notifications. The system sends alert notifications by SMS or email. When monitoring data meets the conditions specified in alert rules, the system sends alert notifications to the corresponding alert contacts.

Procedure

1. *Log on to the Apsara Stack console*.
2. **In the top navigation bar, click  , then choose Administration > CloudMonitor.**
3. **Click the Alert Contact tab to go to the Alert contact sub-page.**
4. **Click Create Contact.**
5. **Configure the alert contact. For more information, see** *Table 1-2: Alert contact configurations*.

Table 1-2: Alert contact configurations

Configuration	Description
Username	The username of the alert contact.

Configuration	Description
Department	The department to which the alert contact belongs. If you select All, the alert contact is a global one.
Project	The project to which the alert contact belongs.
Cell Phone Number	The mobile phone number of the alert contact. The system uses it to send alert notifications to the alert contact by SMS. Make sure that the entered mobile phone number is correct. If the number is changed, update it in time on the platform.
Email	The email address of the alert contact. The system uses it to to send alert notifications to the alert contact by email. Make sure that the entered email address is correct. If the email address is changed, update it in time on the platform.
DingTalk ID	The DingTalk ID of the alert contact.

6. Click OK.

1.6.2.2 Add an alert contact to alert groups

You can add a created alert contact to alert groups for better management.

Prerequisites

- An alert contact is created. For more information, see *Create an alert contact*.
- An alert group is created. For more information, see *Create an alert group*.

Context

An alert contact can be added to multiple alert contact groups.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, click  , then choose Administration > CloudMonitor.
3. Click the Alert Contact tab to go to the Alert Contact sub-tab.
4. Select the alert contact that you want to add to alert groups and then click Add to Alert Group.
5. In the displayed Change Alert Group dialog box, select alert groups and click OK.

1.6.2.3 Query an alert contact

You can query the information and alert groups of alert contacts on the Alert Contact page.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, click  , then choose Administration > CloudMonitor.
3. Click the Alert Contact tab to go to the Alert Contact sub-tab.
4. Select the query condition based on name, cell phone number, email, or DingTalk ID. Enter the keyword in the search bar and then click Search to query the information and alert groups of an alert contact.

1.6.2.4 Modify alert contact information

You can query the information and alert groups of alert contacts on the Alert Contact page.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, click  , then choose Administration > CloudMonitor.
3. Click the Alert Contact tab to go to the Alert Contact sub-tab.
4. Find the alert contact whose information you want to modify. Click the  icon in the Actions column and select Change.
5. In the displayed dialog box, modify the contact information of the alert contact, namely the cell phone number, email address, and DingTalk ID.
6. Click OK.

1.6.2.5 Delete alert contacts

You can delete one or more alert contacts that are no longer in use based on the business requirements.

Procedure

1. *Log on to the Apsara Stack console*.

2. In the top navigation bar, click  , then choose Administration > CloudMonitor.
3. Click the Alert Contact tab to go to the Alert Contact sub-tab.
4. Then, you can:
   - Delete an alert contact

     Find the alert contact to be deleted. Click the icon  in the Actions column and select Remove.
   - Delete multiple alert contacts

     Select multiple alert contacts to be deleted and click Delete Alert Contacts in the upper-right corner.
5. In the displayed dialog box, click OK.

1.6.3 Alert groups

1.6.3.1 Create an alert group

You can create an alert group to classify alert contacts.

Context

An alert group is a group of alert contacts. It contains one or more alert contacts. If an alert is triggered, all the alert contacts in the alert group can receive the alert notification.

When setting an alert rule, you must select an alert group to receive the alert notifications. For each monitoring metric, if the alert threshold is reached, the system sends alert notifications to the members in the alert group according to the configured notification methods.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, click  , then choose Administration > CloudMonitor.
3. Click the Alert Contact tab and click Alert Group sub-page.
4. Click Create Contact Group.

5. Configure the alert group.

Configuration	Description
Group Name	The name of the alert contact group, which must be 2 to 20 characters in length and contain letters, numbers, and underscores (_).
Department	The department to which the alert contact to be added belongs.
Project	The project to which the alert contact to be added belongs.
Description	The description of the alert contact group, which must be 0 to 256 characters in length and can contain letters, numbers, hyphens (-), or underscores (_).
Select Contacts	Add contacts to the alert contact group as follows: Select contacts in the Existing Contacts field and click the  icon to add them to the Selected Contacts. To remove a selected contact, select the contact and then click the  icon. Note: If the contact is not created, create an empty alert group first. Then, *create an alert contact* and *add the alert contact to the alert group*.

6. Click OK.
7. Optional: To remove an alert contact from the alert group, go to the Alert Group page and click Delete at the right of the alert contact.

1.6.3.2 Change alert notification methods

Phone notifications, email notifications, and DingTalk notifications are enabled by default. You can disable the unnecessary notification methods as required.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, click  , then choose Administration > CloudMonitor.
3. Click the Alert Contact tab, click the Alert Contact sub-tab.

4. Find the alert contact whose alert notification methods you want to change. Enable or disable the phone notifications, email notifications, and DingTalk notifications by turning on or off the switches.

1.6.4 Alert rules

1.6.4.1 Create an alert rule

You can create an alert rule for an instance or a bucket to monitor this instance or bucket.

Prerequisites

For Elastic Compute Service (ECS) instances, you must install the monitoring plug-in to collect the metric data at the operating system level.

Follow these steps to install the plug-in:

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, click  , then choose Administration > CloudMonitor.
3. Click the Monitoring tab.
4. In the ECS instance list, find the instance to be monitored. Click the  icon in the Actions column and select Install Plugin.

> **Note:**
> The monitoring chart displays monitoring data 5 to 10 minutes after the monitoring plug-in is installed.

Context

We recommend that you create an alert group before setting an alert rule. You can also create an alert group when you are setting an alert rule. For more information about how to create an alert group, see *Create an alert group*.

Alert rules configured in CloudMonitor are used to monitor the server performance . In this way, you can detect and resolve server problems in time, which guarantees a secure, stable, and effective operation of servers.

Procedure

1. *Log on to the Apsara Stack console*.

2. In the top navigation bar, click  , then choose Administration > CloudMonitor.
3. Click the Monitoring tab.
4. Click a cloud product sub-tab.
5. Find the corresponding instance or bucket. Click the  icon in the Actions column and select Alert Rules to go to the Alert Item page.

   Note:
   You can also use the search function to find a specific instance or bucket and create an alert rule for the instance or bucket.
6. Click Create Alert Rule.
7. Configure the alert rule.

Configuration	Description
Monitor Metric	Select a monitor metric from the drop-down list. For more information about monitor metrics, see *Cloud monitoring metrics*.
Reference Period	Select a reference period from the drop-down list. The reference period is the interval at which data statistics are generated.

Configuration	Description
Condition	Select a condition from the drop-down list. The following conditions are available: · Average: If the average value of all monitoring data collected in a reference period exceeds the threshold, an alert is triggered. · Maximum: If the maximum value of the monitoring data collected in a reference period exceeds the threshold, an alert is triggered. · Minimum: If the minimum value of the monitoring data collected in a reference period exceeds the threshold, an alert is triggered. · Original: If the original value of the monitoring data collected in a reference period exceeds the threshold, an alert is triggered. **Note:** The original value is the condition for Object Storage Service (OSS) product, which refers to the original monitoring data of OSS product.

8. Click Next.
9. Configure the notification object.

   A notification object is an alert contact. For more information about how to configure an alert contact, see *Create an alert contact*.

Configuration	Description
Alert Retries	Select the number of retries before an alert is triggered from the drop-down list. If the value of reference period consecutively exceeds the threshold, an alert is triggered. The system notifies alert contacts only after the number of alert retries is exceeded.
Contact Notification Group	Select a contact notification group. After you set an alert rule for a monitor metric, the system sends an alert notification to the alert contacts if the monitoring data meets conditions configured in the alert rule.

Configuration	Description
Notification Time	Select the notification time, which is a time range during which the system sends alert notifications.

10. Click OK.

1.6.4.2 Create multiple alert rules

You can create the same alert rule for multiple instances or buckets to monitor these instances or buckets.

Prerequisites

For Elastic Compute Service (ECS) instances, you must install the monitoring plug-in to collect the metric data at the operating system level.

Follow these steps to install the plug-in:

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, click , then choose Administration > CloudMonitor.
3. Click the Monitoring tab.
4. In the ECS instance list, find the instance to be monitored. Click the icon in the Actions column and select Install Plugin.

> **Note:**
>
> The monitoring chart displays monitoring data 5 to 10 minutes after the monitoring plug-in is installed.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, click , then choose Administration > CloudMonitor.
3. Click the Monitoring tab.
4. Click a cloud product sub-tab and then select multiple instances or buckets.
5. Click Create Alert Rules in the upper-right corner.

6. Configure the alert rule.

Configuration	Description
Monitor Metric	Select a monitor metric from the drop-down list. For more information about monitor metrics, see *Cloud monitoring metrics*.
Reference Period	Select a reference period from the drop-down list. The reference period is the interval at which data statistics are generated.
Condition	Select a condition from the drop-down list. The following conditions are available: · Average: If the average value of all monitoring data collected in a reference period exceeds the threshold, an alert is triggered. · Maximum: If the maximum value of the monitoring data collected in a reference period exceeds the threshold, an alert is triggered. · Minimum: If the minimum value of the monitoring data collected in a reference period exceeds the threshold, an alert is triggered. · Original: If the original value of the monitoring data collected in a reference period exceeds the threshold, an alert is triggered. **Note:** The original value is the condition for Object Storage Service (OSS) product, which refers to the original monitoring data of OSS product.

7. Click Next.

8. Configure the notification object.

   A notification object is an alert contact. For more information about how to configure an alert contact, see *Create an alert contact*.

Configuration	Description
Alert Retries	Select the number of retries before an alert is triggered from the drop-down list. If the value of reference period consecutively exceeds the threshold, an alert is triggered. The system notifies alert contacts only after the number of alert retries is exceeded.
Contact Notification Group	Select a contact notification group. After you set an alert rule for a monitor metric, the system sends an alert notification to the alert contacts if the monitoring data meets conditions configured in the alert rule.
Notification Time	Select the notification time, which is a time range during which the system sends alert notifications.

9. Click OK.

1.6.4.3 View alert rules

You can view your alert rules on the Alert Rules page after creating an alert rule.

Context

Alert rules are used to display monitoring metrics in the alert rules of CloudMonit or. In this way, you can quickly view monitoring metrics, which guarantees a secure, stable, and effective operation of servers.

Currently, the system provides the alert rules for Elastic Compute Service (ECS), ApsaraDB for RDS, Server Load Balancer (SLB), Object Storage Service (OSS), and KVStore for Redis. The operations used to manage alert rules are similar to each other. Therefore, take the ECS alert rules as an example in this topic.

Procedure

1. *Log on to the Apsara Stack console*.

2. In the top navigation bar, click  , then choose Administration > CloudMonitor.
3. Click the Alert Rules tab.
4. Click the tab of a cloud product, such as ECS. Then the ECS alert rules appear.
5. Enter the ID and name of a monitored resource, select a region, monitoring metric, alert status, and enabled status, and then click Search to query alert rules.

1.6.4.4 View the alert history

After alerts are triggered, you can view the alert history on the Alert Rules page.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, click  , then choose Administration > CloudMonitor.
3. Click the Alert Rules tab.
4. Click the tab of a cloud product.
5. Find the alert rule whose alert history you want to view. Click the  icon in the Actions column and select Alert History.
6. Optional: In the displayed History dialog box, select the start time and end time of the alert and then click Search.
7. View the alert history in the History dialog box.

1.6.4.5 Modify alert rules

You can modify the alert rules of your cloud products on the Alert Rules page.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, click  , then choose Administration > CloudMonitor.
3. Click the Monitoring tab.
4. Click the tab of a cloud product.

5. Find the alert rule to be modified. Click the icon in the Actions column and select Change to modify the alert rule.

   For more information about how to modify alert rules, see *Create an alert rule* to configure alert rules.

1.6.4.6 Pause alert rules

You can pause one or more alert rules as required.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, click , then choose Administration > CloudMonitor.
3. Click the Monitoring tab.
4. Click the tab of a cloud product.
5. Follow these operations:
   - Pause an alert rule.

     Find the alert rule to be paused. Click the icon in the Actions column and select Disable.
   - Pause multiple alert rules.

     Select multiple alert rules to be paused and click Pause in the upper-right corner.
6. In the displayed dialog box, click OK.

   After you pause the alert rules, the system stops to send alert notifications to the corresponding alert contacts.

1.6.4.7 Start alert rules

You can start one or more paused alert rules as required.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, click , then choose Administration > CloudMonitor.

3. Click the Monitoring tab.
4. Click the tab of a cloud product.
5. Follow these operations:
   - Start an alert rule.

     Find the alert rule to be started. Click the  icon in the Actions column and then select Enable.
   - Start multiple alert rules.

     Select multiple alert rules to be started and click Start in the upper-right corner. Then click OK.

1.6.4.8 View alert notification objects

After creating alert rules, you can view the notification object of each alert rule on the Alert Rules page.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, click  , then choose Administration > CloudMonitor.
3. Click the Alert Rules tab.
4. Click the tab of a cloud product.
5. Click the alert contact or alert group in the `Alert Contact` column.

   The system displays the detailed information of alert contacts in the appeared dialog box.

1.6.4.9 Delete alert rules

You can delete one or more alert rules that are no longer in use.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, click  , then choose Administration > CloudMonitor.
3. Click the Monitoring tab.

4. Click the tab of a cloud product.
5. Follow these operations:
   - Delete an alert rule.

     Find the alert rule to be deleted. Click the  icon in the Actions column and select Remove.
   - Delete multiple alert rules.

     Select multiple alert rules to be deleted and click Delete in the upper-right corner. Then, click OK.
6. In the displayed dialog box, click OK.

1.6.5 Configuration of alert notification

1.6.5.1 Configure the email alert notification

The system administrator can configure the email alert notification to allow alert contacts to receive alert notifications by email when alerts are triggered.

Prerequisites

Make sure that the SMTP server address is obtained before you configure the email notification.

To obtain the SMTP server address and port, view the official description of the mailbox system to be configured. Generally, the SMTP server address is in the format of smtp.xxxx.com. For example, the SMTP server address of the 163 mailbox is smtp.163.com.

The system sends email notifications by using the configured email address and email password.

Procedure

1. *Log on to the Apsara Stack console* as a system administrator.
2. In the upper-right corner, click the  icon to go to the System Configuration page.
3. Click the Alert Notification Configuration tab.
4. In the Email Alert Notification Settings section, click Configure.

   The Email Alert Notification Settings dialog box appears.

5. Enter the SMTP server address, email address, and email password, and then select the SMTP server port.
6. Click OK.

   To modify the configurations, click Clear Settings and reconfigure the settings.

1.6.5.2 Configure DingTalk alert notification

The system administrator can configure the DingTalk alert notification to allow alert contacts to receive alert notifications by DingTalk when alerts are triggered.

Context

To send alert notifications by using DingTalk, you must obtain the Corporation ID, Corporation Secret, and Agent ID.

Procedure

1. Obtain the Agent ID.
   a) Log on to oa.dingtalk.com as a DingTalk administrator.
   b) Click Applications and find the Application Base section.
   c) Click the  icon on an application and then select Set.
   d) In the displayed dialog box, obtain the Agent ID.
2. Obtain the Corporation ID, Corporation Secret.
   a) Log on to oa.dingtalk.com as a DingTalk administrator.
   b) Click Applications and find the Create your app section.
   c) Click Open Application to go to the DingTalk developer platform.
   d) In the left-side navigation pane, click Account Management. In the Account Information section, obtain the Corporation ID and Corporation Secret.
3. *Log on to the Apsara Stack console* as a system administrator.
4. In the upper-right corner, click the  icon to go to the System Configuration page.
5. Click the Alert Notification Configuration tab.
6. In the DingTalk Alert Notification Settings section, click Configure.

   The DingTalk Notification Settings dialog box appears.
7. Enter the Corporation ID, Corporation Secret, and Agent ID.

8. Click OK.

   To modify the configurations, click Clear Settings and reconfigure the settings.

1.6.5.3 Configure the SMS alert notification

The system administrator can configure the SMS alert notification to allow alert contacts to receive alert notifications by SMS when alerts are triggered.

Procedure

1. *Log on to the Apsara Stack console* as a system administrator.
2. In the upper-right corner, click the  icon to go to the System Configuration page.
3. Click the Alert Notification Configuration tab.
4. In the SMS Alert Notification Settings section, click Configure.

   The SMS Notification Settings dialog box appears.
5. Enter the Notification URL, AccessKey ID, and AccessKey Secret.
6. Click OK.

   To modify the configurations, click Clear Settings and reconfigure the settings.

1.6.6 View alert information

You can view alert information to have a clear grasp of the running conditions of Elastic Compute Service (ECS), Server Load Balancer (SLB), ApsaraDB for RDS, KVStore for Redis, and Object Storage Service (OSS) and obtain the exception information in time.

Context

Alert information displays the information of alert rules that do not meet the requirements.

Note:

- A maximum of 1000,000 alert records generated within the last 3 months are retained.
- This topic takes ECS alert information as an example to describe how to view the alert information. You can view the alert information of other cloud resources in a similar way.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, click  , then choose Administration > CloudMonitor.
3. Click the Alert Information > ECS tab.
4. You can filter alert information based on the region, monitored resource ID, monitored resource name, monitor metric, start date, and end date. See the following table for the field descriptions of alert information.

   Table 1-3: Field descriptions

Field	Description
Monitored Resource ID/Name	Instance ID or name of the monitored object.
Region	Region in which the monitored object resides.
Monitor Metric	Monitor metric of the monitored object.
Threshold	Threshold of the monitor metric.
Alert Value	Value of the monitor metric when the alert is triggered.
Description	Detailed description of the alert information.
Trigger Status	Alert trigger status, including `Alerts` and `Insufficient Data`.
Start Time	Time when the alert is started.
End Time	Time when the alert is ended.

5. Optional: Click Export to export the current alert information to your local computer as an XLS file.

   The exported file is named *`Alert.xls`* and stored in *`C:\Users\Username\Downloads by default`*.

1.7 Monitoring management

1.7.1 Overview

CloudMonitor provides real-time monitoring, alert, and notification services for resources to protect your products and business.

Currently, CloudMonitor can monitor the metrics of Elastic Compute Service (ECS), Server Load Balancer (SLB), ApsaraDB for RDS, KVStore for Redis, and Object Storage Service (OSS).

You can use the monitoring metrics of cloud products to set alert rules and notification policies to keep up with the running conditions and performance of instance resources for product services. Consider a scale-up in time after receiving an insufficient resource alert.

CloudMonitor has the following functions:

- **Automatic monitoring: The monitoring is automatically started based on your created ECS resources or Auto Scaling groups. You are not required to start it manually or install any plug-ins. You can view the monitoring data of specific instances on the monitoring page after applying for resources.**
- **Flexible alert: You can configure alerts flexibly, such as setting alerts and thresholds for monitoring metrics, pausing, and starting alerts.**
- **Real-time notification: You can set the alert notification to receive notifications by SMS or email in real time. If the status of an alert rule changes, such as alerts are triggered, data is insufficient, or alerts are cleared, the system informs you by SMS or email.**

1.7.2 View dashboard

The Apsara Stack console uses charts to display the usage and monitoring conditions of existing system resources in each region, which helps you learn about the usage of current resources.

Context

> **Note:**
> **Resource types vary with region types. See your dashboard for available resource types.**

Procedure

1. *Log on to the Apsara Stack console*.

2. The system displays the Dashboard page by default. When you are in another page, click Apsara Stack in the upper-left corner to go back to the Dashboard page.
3. Click the region name on the right side of Dashboard to switch regions. You can view the usage overview and alert information of each cloud resource in each region.
   - ECS Instance Loads

     Displays the loads of Elastic Compute Service (ECS) instances in two dimensions, namely CPU utilization and memory usage, and displays the top five instances of CPU loads and memory loads.
   - Regions

     Displays the location information of the current region and other regions on the map. Click other regions to switch region.
   - Available ECS Resources

     Displays the quota data of the top four departments that have the highest CPU quota usage. Click More in the upper-right corner to view available quotas of other departments.
   - Instances by Product

     Displays instance numbers of cloud products in histogram, namely ApsaraDB for Relational Database Service (RDS), Object Storage Service (OSS), ECS, Virtual Private Cloud (VPC), and Server Load Balancer (SLB).
   - Alert information of cloud resources

     Click the corresponding tab of each cloud resource to view its alert information. On each cloud resource tab, click More on the right and then you are redirected to the corresponding alert information page.

     For more information about the alert rules of each cloud resource, see *Cloud monitoring metrics*.

1.7.3 CloudMonitor

1.7.3.1 Overview

CloudMonitor provides real-time monitoring, alert, and notification services for resources to protect your products and business.

You can use the monitoring metrics of cloud products to set alert rules and notification policies to keep up with the running conditions and performance of product instances. Consider a scale-up in time after receiving an insufficient resource alert.

CloudMonitor has the following functions:

- **Automatic monitoring: The monitoring is automatically started based on your created Elastic Compute Service (ECS) resources or Auto Scaling groups. You are not required to start it manually or install any plug-ins. You can view the monitoring data of specific instances on the monitoring page after applying for resources.**
- **Flexible alert: You can configure alerts flexibly, such as setting alerts and thresholds for monitoring metrics, pausing and starting alerts.**
- **Real-time notification: You can set the alert notification to receive notifications by SMS or email in real time. If the status of an alert rule changes, such as alerts are triggered, data is insufficient, or alerts are cleared, the system informs you by SMS or email.**

1.7.3.2 View CloudMonitor overview

You can view the number of instances, number of alert rules, number of alerts, and alert status of all cloud products on the Overview tab of CloudMonitor.

Procedure

1. *Log on to the Apsara Stack console*.
2. **In the top navigation bar, click  , then choose Administration > CloudMonitor.**
3. **On the Overview tab, view the number of instances, number of alert rules, number of alerts, and alert status of all cloud products.**

1.7.3.3 Cloud monitoring metrics

This topic describes the monitoring metrics of each product and the corresponding description.

CloudMonitor tests the service availability based on the monitoring metrics of cloud resources. You can configure alert rules and notification policies for these monitoring metrics to keep up with the running conditions and performance of product instances.

CloudMonitor can monitor resources of Elastic Compute Service (ECS), Server Load Balancer (SLB), ApsaraDB for RDS, Object Storage Service (OSS), and KVStore for Redis. Monitoring metrics supported by each service are described as follows.

Table 1-4: ECS monitoring metrics

Metric name	Metric description	Measured object	Calculation formula	Remarks
CPU Utilization	Measures the CPU utilization (%) of a measured object	ECS instance	CPU utilization of the ECS instance/ Total CPU cores of the ECS instance	None
Memory Utilization	Measures the memory utilization (%) of a measured object	ECS instance	Used memory of the ECS instance /Total memory of the ECS instance	The memory utilization calculated by CloudMonitor does not include cache utilization. Therefore, when you run the `free` or `top` command to query the memory utilization of a Linux server, the result may be inconsistent with the memory utilization displayed in the Apsara Stack console.

Metric name	Metric description	Measured object	Calculation formula	Remarks
Disk I/O Read	Measures the volume of data read from a measured object per second (KB/s)	ECS instance	Total bytes read from the ECS instance disk/ Statistical peirod	For Linux hosts, the disk I/O monitoring data is obtained by using the iostat tool. If your Linux host has no disk I/O data, check if iostat is installed on your machine. If not, Redhat and CentOS users can use yum to install the tool, and Ubuntu and Debian users can use apt-get to install the tool.
Disk I/O Write	Measures the volume of data written to a measured object per second (KB/s)	ECS instance	Total bytes written to the ECS instance disk/ Statistical peirod	None
Disk Utilization	Measures the disk utilization (%) of a measured object	ECS instance	Used capacity of the ECS instance disk/Total capacity of the ECS instance disk	None
Inbound Traffic	Measures the inbound network traffic of a measured object per second (Kbit/s)	ECS instance	None	None

Metric name	Metric description	Measured object	Calculation formula	Remarks
Outbound Traffic	Measures the outbound network traffic of a measured object per second (Kbit /s)	ECS instance	None	If the purchased bandwidth is used up, access fails or requests slow down. On the monitoring chart, eth0 indicates the intranet NIC of the server, and eth1 indicates the Internet NIC of the server.
TCP Connections	Total number of TCP connections set up by the server	ECS instance	None	None
Processes	After you set an alert rule with this monitoring metric, the specified processes are counted and the system displays the total number of these processes.	ECS instance	None	To monitor the running conditions of processes on the server, set an alert rule with this monitoring metric to trigger the alert when the number of processes is unequal to the actual number of processes.
Average Load	A concept in Linux, the average load value of the server	ECS instance		The average load value cannot be greater than 1. If your server has a multi-core processor, the average load value must be divided by the number of CPU cores and the result must be smaller than 1. Generally, if the average load value is greater than 1, processes are queued up and the server slows down.

 **Note:**

For ECS instances, you must install the monitoring plug-in to collect the metric data at the operating system level.

Installation method: Click the Monitoring tab. In the ECS instance list, locate the instance to be monitored. Click the icon in the Actions column and select Install Plugin.

The monitoring chart displays monitoring data 5-10 minutes after the monitoring plug-in is installed.

Table 1-5: RDS monitoring metrics

Metric name	Metric description	Measured object	Calculation formula
CPU Utilization	Measures the CPU utilization (%) of a measured object	ApsaraDB for RDS instance	CPU utilization of the ApsaraDB for RDS instance /Total CPU cores of the ApsaraDB for RDS instance
Memory Utilization	Measures the memory utilization (%) of a measured object	ApsaraDB for RDS instance	Memory usage of the ApsaraDB for RDS instance /Total memory of the ApsaraDB for RDS instance
Disk Utilization	Measures the disk utilization (%) of a measured object	ApsaraDB for RDS instance	None
IOPS Utilization	Measures the number of I/O requests of a measured object per second. Unit: %	ApsaraDB for RDS instance	Number of I/O requests of the ApsaraDB for RDS instance/Statistical period
Connection Utilization	Measures the number of connections between the application and the measured object per second. Unit: %	ApsaraDB for RDS instance	Number of connections between the application and the ApsaraDB for RDS instance per second/ Statistical period

Table 1-6: SLB monitoring metrics

Metric name	Metric description	Measured object	Remarks
Outbound Packets per Second	Number of packets sent by SLB per second	SLB instance	None
Inbound Packets per Second	Number of request packets received by SLB per second	SLB instance	None
Inbound Data	Traffic consumed to access SLB from the external (Kbit/s)	SLB instance	None
Outbound Data	Traffic consumed by SLB to access the external (Kbit/s)	SLB instance	None
Active Port Connections	Number of all connections in the ESTABLISHED status	SLB instance	It can be interpreted as, but cannot be equivalent to, the concurrent connections. This is because a persistent connection transmits multiple file requests simultaneously.
Inactive Port Connections	Number of all TCP connections except connections in the ESTABLISHED status	SLB instance	None
New Port Connections	Number of times the first SYN_SENT status occurs in a TCP three-way handshake during a statistical period	SLB instance	Active Port Connections, Inactive Port Connections, and New Port Connections are all used to measure the number of requests for connecting a client to an SLB instance.

Table 1-7: OSS monitoring metrics

Metric name	Metric description	Measured object
Reads	Measures the number of reads of a measured object	OSS bucket
Internal Server Errors	Measures the number of errors of a measured object	OSS bucket
Public Network Inbound Traffic	Measures the inbound Internet network traffic (bytes) of a measured object per second	OSS bucket
Public Network Outbound Traffic	Measures the outbound Internet network traffic (bytes) of a measured object per second	OSS bucket
Classic Network Inbound Traffic	Measures the inbound intranet network traffic (bytes) of a measured object per second	OSS bucket
Classic Network Outbound Traffic	Measures the outbound intranet network traffic (bytes) of a measured object per second	OSS bucket
Writes	Measures the number of writes of a measured object	OSS bucket
Storage Space Used	Measures the used storage space (bytes) of a measured object	OSS bucket

Table 1-8: KVStore for Redis monitoring metrics

Metric name	Metric description	Measured object	Unit
CPU Utilization	Measures the CPU utilization of a measured object	KVStore for Redis instance	%
Memory Utilization	Measures the proportion of current used memory to the total memory of a measured object	KVStore for Redis instance	%

Metric name	Metric description	Measured object	Unit
Memory Used	Measures the used memory of a measured object	KVStore for Redis instance	Bytes
Connections Used	Measures the total number of connections of the client	KVStore for Redis instance	-
Connection Usage	Measures the proportion of current established connections to the total connections of the measured object	KVStore for Redis instance	%
Input Traffic per Second	Measures the network traffic currently writen per second of a measured object	KVStore for Redis instance	Bytes/s
Output Bandwidth	Measures the network traffic currently read per second of a measured object	KVStore for Redis instance	Bytes/s
Failed Operations	Measures the times of failure of operating a measured object	KVStore for Redis instance	times/s
Write Bandwidth Usage	Measures the proportion of currently writen bandwidth to the total bandwidth of a measured object	KVStore for Redis instance	%
Read Bandwidth Usage	Measures the proportion of currently read bandwidth to the total bandwidth of a measured object	KVStore for Redis instance	%
QPS Usage	Measures the currently used number of QPS of a measured object	KVStore for Redis instance	times/s

1.7.3.4 View monitoring charts

You can view the monitoring chart to learn about the running conditions of each instance or bucket.

Procedure

1. *Log on to the Apsara Stack console*.

2. In the top navigation bar, click  , then choose Administration > CloudMonitor.
3. Click the Monitoring tab.
4. Click the tab of a cloud product.
5. Find the instance or bucket whose monitoring chart you want to view. Click the  icon in the Actions column and select Monitoring Chart.

   You can view the monitoring data of each monitoring metric on the displayed page.

1.7.4 System reports

1.7.4.1 Create a report download task

You must create a report download task on the System Reports page before previewing or downloading reports.

Context

You can create a download task for the following reports:

- **Resource reports**

  A resource report summarizes the current number of instances for cloud products in the Apsara Stack console and the details of each instance, including the region, department, project, and status of the instance.
- **Alert reports**

  An alert report summarizes the alert information of cloud products.
- **Resource usage evaluation reports**

  A resource usage evaluation report summarizes the usage of resources for cloud products. You can view resource usage evaluation reports to learn about the usage of each resource and prevent waste or overload use of resources.
- **Resource monitoring reports**

  A resource monitoring report summarizes the monitoring information of cloud products.

- **Quota reports**

  A quota report summarizes the quota information of cloud products in each department.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click , then choose Administration > System Reports.**
3. **On the System Reports page, click a report tab.**
4. **Configure the filter conditions or evaluation rules based on business requirements and click Create Report Download Task.**
5. **In the displayed Create Report Download Task dialog box, enter a Report name and select the Creation time, Department , and Product. Then click Create.**

   After the report download task is created, you can click Download Center to go to the download center page to view the status of this task.

1.7.4.2 Modify the report name

You can modify the report name on the Download Center page after a download task is created.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click , then choose Administration > System Reports.**
3. **Click the Download Center tab.**
4. **In the task list, find the task whose report name you want to modify. Click the icon in the Actions column and select Change Report Name.**

   Note:
   You can also query the download tasks based on the status or created date of the task to change report names.
5. **In the displayed dialog box, enter the report name and click OK.**

1.7.4.3 Preview and download a report

You can preview and download a report based on the report name and type.

Prerequisites

You can only preview or download a report whose task status is Complete.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click  , then choose Administration > System Reports.**
3. **Click the Download Center tab.**
4. **Find the report to be previewed based on the report name and type. Click the  icon in the Actions column and select Preview.**
5. **Find the report to be downloaded based on the report name and type. Click the  icon in the Actions column and select Download.**
6. **In the displayed dialog box, click OK.**

1.7.4.4 Delete a report download task

You can delete a report download task that is no longer in use.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click  , then choose Administration > System Reports.**
3. **Click the Download Center tab.**
4. **In the task list, find the download task to be deleted. Click the  icon in the Actions column and select Remove.**
5. **In the displayed dialog box, click OK.**

1.7.5 Task center

1.7.5.1 View running tasks

Before a task is finished, the administrator can view the task details in Current Tasks.

Context

You can query tasks quickly by department, task ID, task name, task type, and created time.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click**  **, then choose Administration > System Reports.**
3. **Click the Current Tasks tab.**
4. **Configure the query conditions and then click Search.**

   In the search results, view the task details.

1.7.5.2 View previous tasks

The administrator can view the details of finished tasks in previous tasks.

Context

A previous task is a finished task.

You can query tasks quickly by department, task ID, task name, task type, and created time.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click**  **, then choose Administration > System Reports.**
3. **Click the Previous Tasks tab.**
4. **Configure the query conditions and then click Search.**
5. **Optional: If a task is in** `Error` **status, click Error in the Status column.**

   View the failure details of the task.

1.7.6 Operation logs

1.7.6.1 View logs

You can view logs to learn about the usage conditions of resources on the platform, such as Elastic Compute Service (ECS), ApsaraDB for RDS, and Server Load Balancer (SLB). You can also learn about the running conditions of all function modules on the platform in real time.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click  , then choose Administration > Operation Log to go to the Operation Log page.**
3. **You can query operation logs by username, module, level, instance ID, start date, and end date.**

   For more information about the fields in the search results, see the following table.

   Table 1-9: Field descriptions

Field	Description
Time	Operation time.
Username	Name of the operator.

Field	Description
Module	· ECS: records all actions related to ECS instances, including creating, modifying, deleting, and querying ECS instances. · ApsaraDB for RDS: records all actions related to ApsaraDB for RDS instances, including creating, modifying, deleting, and querying ApsaraDB for RDS instances. · OSS: records all actions related to Object Storage Service (OSS) instances, including creating, modifying, deleting, and querying OSS instances. · Table Store: records all actions related to Table Store instances, including actions of Table Store instances and tables. · SLB: records all actions related to SLB instances, including creating, modifying, deleting, and querying SLB instances. · VPC: records all actions related to Virtual Private Cloud (VPC) instances, including creating, modifying, deleting, and querying VPC instances, and managing EIP, VSwitches, and VRouters. · EIP: records all actions related to Elastic IP Address (EIP) instances , including creating, modifying, deleting, querying, bounding, and unbounding EIP instances. · REDIS: records all actions related to KVStore for Redis instances, including creating, querying, and deleting KVStore for Redis instances . · ESS: records all actions related to Auto Scaling (ESS) instances, including creating, modifying, deleting, querying, and enabling or pausing ESS instances. · ODPS: records all actions related to MaxCompute instances, including creating, querying, updating, and

Field	Description
Operated Object	The instance ID and name of the operation object.
Region	The region in which the operation object resides.
Level	The operation level, including Information, Notification, Warnings, Error, Important, Emergency, Alerts, and Debug.
Actions	The operation type, including logon, logoff, and display.
Details	Brief introduction to the operation purpose.

4. Optional: Click Export to export the current logs to your local computer as an .xls file.

   The exported file is named `log.xls` and stored in `C:\Users\Username\Downloads`.

1.7.6.2 Delete logs

The administrator can delete logs that are within a specific time period if the logs are no longer in use.

Context

Notice:
Logs cannot be restored after being deleted, so proceed with caution.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click  , then choose Administration > Operation Log to go to the Operation Log page.
3. Configure the query conditions and then click Search.
4. Click Delete Log, and select a specific time period to delete logs.
5. Click OK.

1.8 Resource Access Management

1.8.1 Overview

Resource Access Management (RAM) is a resource access control service provided by Alibaba Cloud Apsara Stack.

RAM allows you to manage users (such as employees, systems, and applications) in a centralized way and control permissions to allow users to access specific resources under your name.

RAM has the following two functions:

- **RAM role**

  You must create a corresponding RAM role to authorize cloud services in a level-1 department to use or view other resources of the current department. This role contains the operations that cloud services can perform on resources.

  Only the system administrator and level-1 department manager can create RAM roles.

- **RAM user**

  To allow multiple users to use cloud resources in the same department, you can create multiple RAM users (users with the Independent Software Vendors (ISV

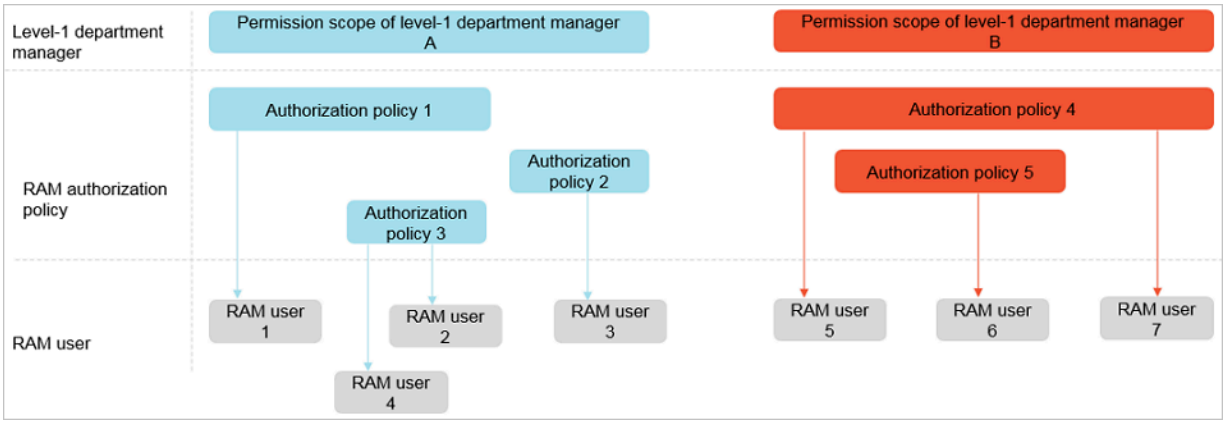
) Account role) in the department to allow multiple RAM users to have different permissions to the same cloud resource.

You can create RAM authorization policies to grant different permissions to different RAM users.

RAM users can be considered as sub-accounts of the creator. A RAM user comes from the creator account based on the authorization policies, and the permission scope of a RAM user is smaller than or equal to that of the creator account.

The RAM users are created by the system administrator or level-1 department manager.

Figure 1-3: Authorization of RAM users



1.8.2 RAM roles

1.8.2.1 View a role policy

You can view a role policy to learn about the detailed authorization of a role.

Prerequisites

A RAM role is created. For more information, see *Create a RAM role*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click , then choose Compute, Storage & Networking > Resource Access Management.
3. On the RAM Role tab, find the role whose policy you want to view. Click the icon in the Actions column and select View Details.

4. Click the Role Policy tab.
5. In the policy list, find the policy that you want to view. Click the  icon in the Actions column and select View Details to view the policy details, namely the name, description, type, and contents of the policy.

1.8.2.2 Create a RAM role

To allow a cloud service to access other cloud resources, you must create the corresponding RAM role of this cloud service in the level-1 department.

Procedure

1. *Log on to the Apsara Stack console* as a system administrator or level-1 department manager.
2. In the top navigation bar, click  , and then choose Compute, Storage & Networking > Resource Access Management.
3. Click the RAM Roles tab.
4. In the upper-right corner, click Create RAM Role.
5. Select the level-1 department, region, and cloud service to be authorized. Then, click Create to complete the authorization.

   The created RAM role appears in the RAM role list.

   For more information about the RAM roles that can be created and their role names, see *Table 1-10: Mapping between RAM roles and services*.

   Table 1-10: Mapping between RAM roles and services

Role name	Service	Role description
AliyunCloudFirewallAccessingECSRole	Cloud Firewall	Used to grant Cloud Firewall to use this role to access Elastic Compute Service (ECS).
AliyunECSImageExportDefaultRole	ECS	Used to grant ECS to use this role to export images.
AliyunECSImageImportDefaultRole		Used to grant ECS to use this role to import images.

Role name	Service	Role description
AliyunCSDefaultRole	Container Service	Used to grant Container Service to use this role to access resources of other cloud products during cluster operations.
AliyunCSClusterRole		Used to grant Container Service to use this role to access resources of other cloud products when applications are running.
AliyunESSDefaultRole	Auto Scaling	Used to grant Auto Scaling to use this role to access resources of other cloud products.
AliyunEMRDefaultRole	E-MapReduce	Used to grant E-MapReduce to use this role to access resources of other cloud products.
AliyunEMRECSDefaultRole		Used to grant E-MapReduce jobs to use this role to access your cloud resources.
AliyunDTSDefaultRole	Data Transmission Service (DTS)	Used to grant DTS to use this role to access resources of other cloud products.
AliyunStreamDefaultRole	StreamCompute	Used to grant Realtime Compute to use this role to access resources of other cloud products.

For example, select A as the Region, B as the Department, and ECS as the Service to create a RAM role. After the RAM role is created, ECS in region A, department B and its sub-departments can use the created RAM role to access resources of other cloud products in region A, department B and its sub-departments.

1.8.2.3 View role details

You can view the role details to learn about the name, description, created time, and global resource descriptor of the role.

Prerequisites

A RAM role is created. For more information, see *Create a RAM role*.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click**  **, then choose Compute, Storage & Networking > Resource Access Management.**
3. **On the RAM Role tab, find the role that you want to view. Click the**  **icon in the Actions column and select View Details.**

   On the Role Details tab, view the role details, namely the name, description, created time, and global resource descriptor of the role.

1.8.3 RAM users

1.8.3.1 Create a RAM user

To allow multiple users to use cloud resources in the same department, you can create multiple RAM users in the department.

Prerequisites

Before creating a RAM user, make sure that a department is created. For more information, see *Create a department*.

Context

RAM users are the operation objects who have specific access permissions to cloud resources. Currently, this function only applies to Object Storage Service (OSS) and Table Store.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click**  **, then choose User Center > User Management.**

3. Click the Users tab.
4. Click Create. The Add User dialog box appears.
5. Configure the user information. Parts of the configurations are as shown in *Table 1-11: RAM user configurations*.

   Table 1-11: RAM user configurations

Configuration	Description
Username	The name must be 2 to 30 characters in length and can contain letters in uppercase or lowercase , numbers, hyphens (-), underscores (_), periods (.), and at signs (@). It must start with a letter or a number.
Display Name	The name must be 2 to 30 characters in length and only contain letters.
Role	Select Independent Software Vendors (ISV) Account. If you select other roles, the RAM user cannot be queried.
Department	Select the department to which the user belongs.
Logon Policy	The logon policy restricts the time period and IP addresses for the user to log on. By default, the default policy is automatically bound to newly created users. **Note:** The default policy does not restrict the time period and IP addresses for users to log on. To restrict the logon time and IP addresses of a user, you can modify the user's logon policy or create a logon policy for the user. For more information, see *Create a logon policy*.

6. Click OK to create a RAM user.

1.8.3.2 View RAM user details

The administrator can view the details of a RAM user created by the administrator.

Prerequisites

A RAM user is created. For more information, see *Create a RAM user*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click  , and then choose Compute, Storage & Networking > Resource Access Management.
3. Click the RAM Users tab.
4. Optional: Select the department where the RAM user resides and enter the username. Click Search to query the RAM user.
5. Find the RAM user that you want to view. Click the  icon in the Actions column and select View Details.

   On the User Details page, view the detailed information of the RAM user, including the username, UID, contact information, and created time.

1.8.3.3 Modify the description of a RAM user

The administrator can modify the description of a created RAM user for better management.

Prerequisites

A RAM user is created. For more information, see *Create a RAM user*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click  , then choose Compute, Storage & Networking > Resource Access Management.
3. Click the RAM Users tab.
4. Optional: Select the department where the RAM user resides and enter the username. Click Search to query the RAM user.
5. Find the RAM user whose description you want to modify. Click the  icon in the Actions column and select Change.
6. In the displayed dialog box, enter the description and then click OK.

1.8.3.4 Grant permissions to a RAM user

The administrator can bind created authorization policies to a RAM user based on the business requirements.

Prerequisites

- **A RAM user is created. For more information, see** *Create a RAM user*.
- **A RAM authorization policy is created. For more information, see** *Table 1-15: Methods supported by operation types*.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click**  **, and then choose Compute, Storage & Networking > Resource Access Management.**
3. **Click the RAM Users tab.**
4. **Optional: Select the department where the RAM user resides and enter the username. Click Search to query the RAM user.**
5. **Find the RAM user that you want to grant permissions. Click the**  **icon in the Actions column and select Authorize.**
6. **In the displayed dialog box, select the authorization policies from the Available Authorization Policies and then click**  **to add them to the Authorization Policies Selected.**
7. **Click OK.**

1.8.4 RAM authorization policies

1.8.4.1 Create a RAM authorization policy

You can create authorization policies as required to grant these authorization policies to RAM users.

Context

RAM authorization policies are the implementations of RAM user permissions. RAM users obtain the permissions by binding RAM authorization policies. Currently, this function only applies to Object Storage Service (OSS) and Table Store.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click  , then choose Compute, Storage & Networking > Resource Access Management.**
3. **Click the RAM Authorization Policies tab.**
4. **Click Create Authorization Policy. The Set Basic Information step appears.**
5. **Configure the basic information of the authorization policy.**

   For more information about the configurations, see *Table 1-12: Authorization policy configurations*.

   Table 1-12: Authorization policy configurations

Configuration	Description
Policy Name	The RAM authorization policy name, which must be 1 to 128 characters in length and can contain letters, numbers, and hyphens (-), but must not start with dtrole.
Region	The region to which the RAM authorization policy belongs.
Department	The department to which the RAM authorization policy belongs.
Project	The project to which the RAM authorization policy belongs.
Product	The product of the RAM authorization policy. Currently, only OSS and Table Store are supported.
Policy Description	The description of the RAM authorization policy.

6. **Click Next to go to the Add Rules and Conditions page.**
7. **Configure the contents of the authorization policy.**

   For more information about the configurations, see *Table 1-13: Content configurations*.

   Table 1-13: Content configurations

Configuration	Description
Effect	The authorization effectiveness includes `Allow` and `Deny`.

Configuration	Description
Action	The operations performed on specific resources. For example, the access policy allows user A to perform the GetBucket operation on the resource SampleBucket. The operation is `GetBucket`.
Resource	The resource is the specific object that is authorized. For example, the access policy allows user A to perform the GetBucket operation on the resource SampleBucket. The resource is `SampleBucket`.
Condition	Composed of one or more condition clauses. A condition clause is composed of the operation type, keyword, and value.
Keywords	The keyword of the condition. For more information, see *Table 1-14: Keywords*.
Operation Type	The operation type of the condition. The following operation types are supported: · String · CurrentTime · Date and time · Boolean · IP address For the methods that each operation type supports, see *Table 1-15: Methods supported by operation types*.

Configuration	Description
Value	The value of the condition.

Table 1-14: Keywords

Keyword	Type	Description
acs:CurrentTime	Date and time	The time when the Web server receives the request, which is in the format of ISO 8601. For example, 2012-11-11T23:59:59Z.
acs:SecureTransport	Boolean	Whether the secure channel (for example, HTTPS) is used for sending the request.
acs:SourceIp	IP address	The client IP address when the request is sent.
oss:Delimiter	String	The delimiter used by OSS to divide object names into groups.
oss:Prefix	String	The prefix of the OSS object name.

Table 1-15: Methods supported by operation types

Operation type	Method
String	· StringEquals · StringNotEquals · StringEqualsIgnoreCase · StringNotEqualsIgnoreCase · StringLike · StringNotLike

Operation type	Method
CurrentTime	· DateEquals · DateNotEquals · DateLessThan · DateLessThanEquals · DateGreaterThan · DateGreaterThanEquals
Date and time	· DateEquals · DateNotEquals · DateLessThan · DateLessThanEquals · DateGreaterThan · DateGreaterThanEquals
Boolean	Bool
IP address	· IpAddress · NotIpAddress

8. Click Add Condition and then click Create Rules.

   The authorization policy details are automatically generated in the Authorization Policy field.

9. Click OK to create the RAM authorization policy.

1.8.4.2 View RAM authorization policy details

You can view the RAM authorization policy details to learn about the name, department, region, and created time of that authorization policy.

Prerequisites

A RAM authorization policy is created. For more information, see *Create a RAM authorization policy*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click  , then choose Compute, Storage & Networking > Resource Access Management.
3. Click the RAM Authorization Policies tab.

4. Optional: Select the department and region where the RAM authorization policy resides and enter the authorization policy name. Click Search to query the RAM authorization policy.
5. Find the RAM authorization policy that you want to view. Click the icon in the Actions column and select View Details.

   On the Policy Details page, view the authorization policy details, namely the name, type, version number, created time, description, and the number of times that the authorization policy is referenced.

1.8.4.3 Delete a RAM authorization policy

You can delete a RAM authorization policy that is no longer in use.

Prerequisites

A RAM authorization policy is created. For more information, see *Create a RAM authorization policy*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click , then choose Compute, Storage & Networking > Resource Access Management.
3. Click the RAM Authorization Policies tab.
4. Optional: Select the department and region where the RAM authorization policy resides and enter the authorization policy name. Click Search to query the RAM authorization policy.
5. Find the RAM authorization policy to be deleted. Click the icon in the Actions column and select Delete.
6. In the displayed dialog box, click OK.

1.9 System maintenance

1.9.1 Department management

1.9.1.1 Create a department

You can create a department to store projects and resources in the projects.

Context

After the Apsara Stack console is deployed, a root department is created by default. You can create departments under the root department. The department s appear hierarchically and you can add sub-departments under each level of the department.

Departments added under the root department are level-1 departments and departments added under the level-1 departments are level-2 departments. Other departments are added in a similar way. In the Apsara Stack console, the sub-departments of a department refer to departments of all levels under the department. You can create at most five levels of departments.

Departments reflect the tree structure of an enterprise or business unit. A user can only belong to one department.

You can create a department under an existing department. The created department is a sub-department of the existing department.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click  , then choose User Center > Department Management.**
3. **Select a department and click Add Department.**
4. **In the displayed Add Department dialog box, enter the department name.**

   The name must be 2 to 50 characters in length and can contain letters and numbers.
5. **Click OK.**

1.9.1.2 Modify the department name

If the department information is changed, the administrator can modify the department name.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click  , then choose User Center > Department Management.**
3. **Select the department whose name you want to modify and click Change Department.**
4. **In the displayed dialog box, modify the department name and click OK.**

1.9.1.3 View projects of a department

You can view projects of a department to learn about the project information in the department.

Context

Departments reflect the tree structure of an enterprise or business unit. A department can have multiple projects.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click  , then choose User Center > Department Management.**
3. **Select the department that you want to view.**

   Projects of this department appear in the right list.

1.9.1.4 Obtain the AccessKey of a department

The administrator can obtain the department AccessKey.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click  , then choose User Center > Department Management.**

3. Select a department and click Get AccessKey to obtain the account name, AccessKey, and PrimaryKey of the department.

> **Note:**
> The system automatically allocates the Apsara Stack account name, AccessKey, and PrimaryKey to the level-1 department. The sub-departments use the same account name, AccessKey, and PrimaryKey as their level-1 department.

1.9.1.5 Delete a department

The administrator can delete a department that is no longer in use.

Prerequisites

> **Notice:**
> Make sure that the department to be deleted does not contain any users, projects, or sub-departments. Otherwise, the department cannot be deleted.

Procedure

1. *Log on to the Apsara Stack console* as the administrator.
2. In the top navigation bar, click  , then choose User Center > Department Management.
3. Select the department to be deleted and click Delete.

1.9.2 Project management

1.9.2.1 Create a project

You must create a project before applying for resources.

Prerequisites

Make sure that a department is created before creating a project. For more information, see *Create a department*.

Context

You can create at most 20 projects under each level-1 department.

Procedure

1. *Log on to the Apsara Stack console* as the administrator.

2. In the top navigation bar, click  , then choose User Center > Project Management.
3. Click Add Projects.
4. In the displayed Add Projects dialog box, enter the project name and select a department to which the project belongs.
5. Click OK.

1.9.2.2 Add a project member

You can add a member to a project to allow the member to use the resources of the project.

Prerequisites

Before adding a project member, make sure that:

- A project is created. For more information, see *Create a project*.
- A user is created. For more information, see *Create a user*.

Context

The members of a project have the permissions to use resources of the project.

If you delete resources from a project, it does not affect the members of the project . Similarly, if you delete members from a project, it does not affect the resources of the project.

You can delete the project members that are no longer in use. A deleted project member cannot access the resources of the project.

Procedure

1. *Log on to the Apsara Stack console* as the administrator.
2. In the top navigation bar, click  , then choose User Center > Project Management.
3. Find the project that you want to add members. Click the icon in the Actions column  > View details.
4. Click the Project Members tab.
5. Click Add Members.

6. In the displayed Add Project Members dialog box, select a department and the corresponding project members.
7. Click OK.

   To remove one or more members from the project, follow these steps:

   a. Select one or more members and click Delete Members.
   b. In the displayed Delete Members dialog box, select Yes.
   c. Click OK.

Result

The project member is added. You can view information about this member in the project member list.

1.9.2.3 Modify project information

If the project information is changed, the administrator can modify the name, description, and additional information of the project.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click  , then choose User Center > Project Management.
3. Find the project that you want to modify. Click the  icon in the Actions column and select Change Project Information.
4. In the displayed dialog box, modify information such as the project name and click OK.

1.9.2.4 View project details

You can view project details to learn about the basic information of a project, including the name, ID, department, created time, and headcount.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click  , then choose User Center > Project Management.

3. Find the project that you want to view. Click the  icon in the Actions column and select View Details.
4. On the Project Details page, view the project details.

1.9.2.5 View project members

To use resources of a project, you must be a member of the project. Check if you are in the member list of the project.

Context

The members of a project have the permissions to use resources of the project.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click  , then choose User Center > Project Management.
3. Find the project. Click the  icon in the Actions column and select View Details.
4. On the Project Details page, click the Project Members tab. You can view all the members of the project and their contact information.

1.9.2.6 View resource information of a project

If you want to use certain cloud resources, you can view the resource information of a project in the project resource list.

Context

The project resource list displays all cloud resources of the project.

Procedure

1. *Log on to the Apsara Stack console* as the administrator.
2. In the top navigation bar, click  , then choose User Center > Project Management.
3. Find the project. Click the  icon in the Actions column and select View Details.
4. On the Project Details page, click the Project Resources tab.
5. On the Project Resources tab, view all cloud resources of the project.

6. Click the tab of a cloud product.
7. Find the resource. Click the icon in the Actions column and select View Details to view the resource details.

1.9.2.7 Release resources

The administrator can release the resources that are no longer in use in a project.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click , then choose User Center > Project Management.
3. Find the project. Click the icon in the Actions column and select View Details.
4. Select the Project Resources tab and then:
   - Release a single resource.

     Click the tab of a cloud product. Find the resource to be released. Click the icon in the Actions column and select Release Resources. In the displayed dialog box, click OK.
   - Release multiple resources.

     Click the tab of a cloud product. Select multiple resources to be released, and then click Delete in the upper-right corner.

1.9.2.8 Delete a project

If a project is finished or changed, the administrator can delete the project that is no longer in use.

Prerequisites

Make sure that the project to be deleted does not contain any resources or project members.

Context

Notice:
If a project contains resources or project members, it cannot be deleted.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click  , then choose User Center > Project Management.
3. Find the project to be deleted. Click the  icon in the Actions column and select Delete Project.
4. Click OK.

1.9.3 Role management

1.9.3.1 Add a custom role

You can add custom roles in the Apsara Stack console to better assign permissions to users.

Context

A role is a collection of access permissions. Each role corresponds to a range of permissions. A user can have multiple roles, which means that this user has all the permissions defined in these roles. You can use a role to grant the same permissions to a group of users.

To meet the requirements that different features can be managed by professionals, the platform supports users to select permissions as required.

Before adding a custom role, note that the total number of custom and default roles cannot exceed 20.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click  , then choose User Center > Role Management.

3. Click Create. The Create Roles dialog box appears.

   Complete the configurations. For more information, see *Table 1-16: Role configurations*.

   Table 1-16: Role configurations

Configuration	Description
Role Name	The name of the role, which must be 1 to 15 characters in length and contain letters or numbers.
Description	(Optional) The description of the role, which must be 1 to 100 characters in length and contain letters, numbers, commas (,), semicolons (;), and underscores (_).
Permission Scope	• Department The permissions apply to all departments of the corresponding modules. • Current Department/Subdepartments The permissions apply to the department to which the user belongs and its subdepartments. • Project The permissions apply to the projects to which the user is added.
Select Permissions	Specify the operation permissions to cloud products. The system selects dependent permissions automatically that are related to the permissions selected by users. If the dependent permissions are removed, users cannot perform the current selection of permissions.

4. Click OK.

1.9.3.2 View role details

You can view permissions of a role on the Role Management page.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.

2. In the top navigation bar, click  , then choose User Center > Role Management.
3. Find the role whose permissions you want to view. Click the  icon in the View Permissions column. View the permissions of this role in the displayed dialog box.

1.9.3.3 Modify a custom role

The administrator can modify the description and permissions of a custom role.

Context

 **Note:**
System default roles cannot be modified.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click**  **, then choose User Center > Role Management.**
3. **Find the role to be modified. Click the**  **icon in the Actions column and select Change.**
4. **In the displayed dialog box, modify the description, permission scope, and permission list of the role.**
5. **Click OK.**

1.9.3.4 Delete a custom role

The administrator can delete a custom role that is no longer in use to manage roles better.

Context

 **Note:**
System default roles cannot be deleted.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**

2. In the top navigation bar, click  , then choose User Center > Role Management.
3. Find the role to be deleted. Click the  icon in the Actions column and select Delete.

   The Confirm Deletion dialog box appears.
4. Click OK.

1.9.4 User management

1.9.4.1 Create a user

The administrator can create a user and assign different roles to the user to meet user's access control requirements on the system.

Prerequisites

Before creating a user, make sure that:

- A department is created. For more information, see *Create a department*.
- A custom role is created if you want to customize the role. For more information, see *Add a custom role*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click  , then choose User Center > User Management.
3. Click the Users tab.
4. Click Create. The Add User dialog box appears.
5. Configure the user information.

Configuration	Description
Username	The cloud platform account name of the user. The name must be 2 to 30 characters in length and can contain letters, numbers, hyphens (-), underscores (_), periods (.), and at signs (@). It must start with a letter or number.

Configuration	Description
Display Name	The name must be 2 to 30 characters in length and can contain letters, numbers, hyphens (-), underscores (_), periods (.), and at signs (@).
Department	Select the department to which the user belongs.
Role	Select a role for the user.
Logon Policy	Select a logon policy for the user. It restricts the time period and IP addresses for the user to log on. By default, the default policy is automatically bound to newly created users. **Note:** The default policy does not restrict the time period and IP addresses for users to log on. To restrict the logon time and IP addresses of a user, you can modify the user's logon policy or create a logon policy for the user. For more information, see *Create a logon policy*.
Cell Phone Number	The mobile phone number of the user. It is used to notify the user of resource applications and usage by SMS. **Note:** Make sure that the entered mobile phone number is correct.
Landline	(Optional) The landline number of the user. It must be 4 to 20 characters in length and can contain numbers (0 to 9) and hyphens (-).
Email	The email address of the user. The system uses it to notify the user of resource applications and usage by email. Make sure that the entered email address is correct. **Note:** If the email address is changed, update it in time on the platform.

For the relationships among departments, users, and roles, see *Configuration of system initialization*.

6. Click OK.

1.9.4.2 View basic information of a user

You can view the basic information of a user to learn about the department, role, and contact information of the user.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click  , then choose User Center > User Management.**
3. **Click the Users tab.**
4. **Find the user whose basic information you want to view. Click the  icon in the Actions column and select User Information to view the basic information of the user.**

1.9.4.3 Modify user information

If the user information is changed, you can modify the display name and contact information of the user.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click the  , then choose User Center > User Management.**
3. **Click the Users tab.**
4. **Find the user to be changed. Click the  icon in the Actions column and select Change.**
5. **In the displayed Change User dialog box, modify the display name and contact information of the user.**

   For more information about how to modify the personal information, see *Modify personal information*.

1.9.4.4 Change the logon policy of a user

For better management, the administrator can change the logon policy of a user to change the permitted logon time and IP addresses for the user.

Prerequisites

A logon policy is created. For more information, see *Create a logon policy*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click  , then choose User Center > User Management.
3. Click the Users tab.
4. Find the user. Click the  icon in the Actions column and select Assign Logon Policy.
5. In the displayed Assign Logon Policy dialog box, select the corresponding logon policy.
6. Click OK.

   After the logon policy of the user is changed, the user is limited by the new policy .

   If the user does not want to be limited by the bound logon policy, the user must submit an application to the administrator. After approving the application, the administrator binds a logon policy that meets the user's requirements to the user .

1.9.4.5 Change user roles

You can change user roles by adding, changing, or deleting roles for a user.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click  , then choose User Center > User Management.
3. Click the Users tab.
4. Find the user to be changed. Click the  icon in the Actions column and select Authorize.
5. In the `Role` field, add, change, or delete roles for the user as required.
6. Click OK.

1.9.4.6 Authorize third-party access

To call APIs of the Apsara Stack console, you must authorize third-party access to obtain the third-party AccessKey.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click , then choose User Center > User Management.**
3. **Click the Users tab.**
4. **Find the user. Click the icon in the Actions column and select Authorize Third-Party Access.**
5. **In the displayed dialog box, click Authorize.**

   Note:
   Authorize Third-Party Access is enabled by default. You can click Recreate an AccessKey or Remove the AccessKey in the displayed dialog box.

   To view the third-party AccessKey, see *View third-party AccessKey*.

1.9.4.7 Reset logon password

The system administrator can reset the logon passwords for users if they forget their logon passwords.

Prerequisites

Only users who have the write permission to user management and project management can reset the logon passwords.

Procedure

1. *Log on to the Apsara Stack console* **as an administrator.**
2. **In the top navigation bar, click , then choose User Center > User Management.**
3. **Click the Users tab.**
4. **Find the user. Click the icon in the Actions column and select User Information.**

5. On the User Information page, click Reset Password. The system automatically generates a new password and sends the new password to the user by SMS.
6. In the displayed Reset Password dialog box,
   - Click Reset Only. The user password is reset to the initial password.
   - Click Reset and Download. The user password is reset to the initial password and locally downloaded in the TXT format.

1.9.4.8 Export initial password

If Reset and Download is selected when the logon password is reset, the administrator can export the initial user password and notify the user of the corresponding logon password orally.

Prerequisites

The password is reset. For more information, see *Reset logon password*.

Procedure

1. *Log on to the Apsara Stack* as an administrator.
2. In the top navigation bar, click  , then choose User Center > User Management.
3. Click the Users tab.
4. Select the user whose initial password you want to export and click Export Initial User Password.

   The password file `UserInitPassword.txt` is generated.

1.9.4.9 Enable and disable a user

To prevent a user from logging on to the Apsara Stack console, you can disable the user. A disabled user must be activated before logging on to the Apsara Stack console.

Context

A user is activated by default after being created.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.

2. In the top navigation bar, click  , then choose User Center > User Management.
3. Click the Users tab.
4. Follow these operations:
   - Find an `Enabled` user. Click the  icon in the Actions column and select Disable to disable this user.
   - Find a `Disabled` user. Click the  icon in the Actions column and select Enable to enable this user.

1.9.4.10 Delete a user

The administrator can delete a user based on business requirements.

Prerequisites

The user is removed from all projects. For more information, see *Add a project member*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click  , then choose User Center > User Management.
3. Click the Users tab.
4. Find the user. Click the  icon in the Actions column and select Delete.
5. In the displayed dialog box, click OK.

   The deleted user still exists in the database, but does not belong to any department or have any role, and cannot log on to the Apsara Stack console.

1.9.4.11 Restore a user

After a user is deleted, the administrator can locate and restore the user in the Deleted Users list.

Context

Except for the department and role information, other basic information and the logon password of a restored user are the same as those before the user was deleted .

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click  , then choose User Center > User Management.
3. Click the Deleted Users tab.
4. Find the user to be restored. Click the  icon in the Actions column and select Restore.

   The Restore User dialog box appears.
5. Select a department and a role and then click OK.

1.9.5 Logon policy management

1.9.5.1 Create a logon policy

The administrator can configure logon policies to control the logon address and time of users.

Context

A logon policy is used to control the logon time and IP address of users. After binding a logon policy to a user, the user can only log on to the Apsara Stack console within the time period and IP addresses configured in the logon policy.

A default logon policy is automatically generated when the Apsara Stack console provides services. This policy does not have any limits on the logon time and IP address, and cannot be deleted.

With the logon policies configured, users can access the Apsara Stack console at the permitted time and from permitted IP addresses. This improves the security of the console.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click  , then choose User Center > Logon Policy Management.
3. Click Create Policy.

4. In the displayed Configure Policy dialog box, enter the policy name, permitted logon/logoff time, and permitted logon IP address.

   Table 1-17: Logon policy configurations

Configuration	Description
Policy Name	The name must be 1 to 15 characters in length and contain letters or numbers, but must not be the same as an existing policy name. The name of the default policy cannot be changed.
Logon/Logoff Time	The permitted logon and logoff time consist of a time period . After being configured, users can only log on to the Apsara Stack console during the specified time period.
Client IP Addresses	The permitted logon address is an IP address Classless Inter-Domain Routing (CIDR) block. After being configured, users can only log on to the Apsara Stack console from the IP addresses within the specified CIDR block.

5. Click OK.
6. Optional: Bind a logon policy to a user. For more information, see *Change the logon policy of a user*.

   Note:

   - After binding a logon policy to a user, this user can only log on to the Apsara Stack console at the permitted time and from permitted IP addresses configured in the policy.
   - If the user does not want to be limited by the bound logon policy, the user must submit an application to the administrator. After approving the application, the administrator binds a logon policy that meets the user's requirements to the user.

What's next

After creating a logon policy, you can modify or delete the existing logon policy.

- Find the logon policy to be modified. Click the icon in the Actions column and select Change to modify the policy.

- Find the logon policy to be deleted. Click the  icon in the Actions column and select Delete to delete the policy.

> **Note:**
> You cannot delete the default logon policy.

1.9.5.2 View a logon policy

You can view a logon policy to learn about the permitted logon time and IP addresses of a user.

Context

A default logon policy is automatically generated when the Apsara Stack console provides services. This policy does not have any limits on the logon time and IP addresses.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, click  , then choose User Center > Logon Policy Management.
3. Optional: Enter the policy name in the search bar and click Search.

   The search result appears.
4. View the corresponding logon policy, namely the permitted logon time and IP addresses of users.

1.9.5.3 Bind a logon policy to multiple users

You can bind the same logon policy to different users. Users are limited by the logon policy when logging on to the Apsara Stack console.

Prerequisites

- Users are created. For more information about how to create users, see *Create a user*.
- A logon policy is created. For more information about how to create a logon policy, see *Create a logon policy*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.

2. In the top navigation bar, click  , then choose User Center > User Management.
3. Click the Users tab.
4. Select multiple users and click Assign Logon Policy to bind a logon policy to multiple users. Then, click OK.

1.9.6 System configuration

1.9.6.1 Configure the storage path for attachments

You can specify the storage path for uploaded attachments by configuring the system Object Storage Service (OSS).

Prerequisites

Before configuring the system OSS, select an OSS bucket as the system OSS and obtain the AccessKey ID and AccessKey Secret. AccessKey ID and AccessKey Secret are used to identify a visitor. The system uses AccessKey ID and AccessKey Secret to access OSS.

Obtain the AccessKey ID and AccessKey Secret as follows:

1. *Log on to the Apsara Stack console* as a system administrator.
2. In the top navigation bar, click  , then choose Compute, Storage & Networking > Object Storage Service. In the bucket list, view the region and department to which the bucket belongs.
3. In the top navigation bar, click  , then choose User Center > Department Management. In the department tree, find the region and department of the bucket. Select the department and then click Get AccessKey.

Context

By default, the storage path for attachments is not configured in the Apsara Stack console, and no attachment upload function is available. Configure the system OSS to specify the storage path for attachments to implement the high-reliability storage of large numbers of attachments.

Procedure

1. *Log on to the Apsara Stack console* as a system administrator.
2. In the upper-right corner, click the  icon to go to the System Configuration page.
3. Click the Storage Configuration tab.
4. Set Storage Type to `OSS`.
5. Configure the system OSS.

   For more information about the configurations, see *Table 1-18: System OSS configurations*.

   Table 1-18: System OSS configurations

Configuration	Description
OSS Endpoint	The endpoint address of OSS. Obtain the endpoint by viewing the bucket details.
Bucket Name	The name of the bucket.
AccessKey ID, AccessKey Secret	The keys used to access OSS. AccessKey ID is used to identify a user, and AccessKey Secret is a key used to authenticate a user.

6. Click Save.
7. Click Test Connection.

   To modify the OSS configuration, click Reset and configure the system OSS again.

1.9.6.2 Configure the access control

Configure the number of terminals that a system account can log on to limit the number of browsers or hosts that the account can log on simultaneously.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the upper-right corner of the page, click  to go to the System Configuration page.

3. Click the Access Control Configuration tab, and then click Configuration Details to modify the maximum allowed logons.

   Note:

   - Enter an integer from 0 to 9999.
   - Enter -1 if no limit is required.

1.9.6.3 Configure the ECS startup

On the Resource Notification Configuration tab, configure whether to automatically start the ECS instance after it is created.

Procedure

1. *Log on to the Apsara Stack console* as a system administrator.
2. In the upper-right corner, click the  icon to go to the System Configuration page.
3. Click the Resource Notification Configuration tab.
4. In the ECS Startup Configuration section, select the `Automatically start the ECS instance after it is created` check box.
5. Click Save.

   A system prompt appears, indicating the instance has been configured.

1.9.7 Configure the theme

You can change the theme of system as required.

Procedure

1. *Log on to the Apsara Stack console* as a system administrator.
2. Click the avatar in the upper-right corner and select Theme.
3. Go to the Theme Configuration page.
4. Select the theme type as required, and then click Save.

1.10 Personal information management

1.10.1 Modify personal information

If your personal information is changed, you can modify the basic information of the personal information.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the upper-right corner, click your avatar and select Personal Information.
3. Click Change at the right of the item.
4. Modify the information.
5. Click Save.

1.10.2 View AccessKey of your personal account

To guarantee the security of cloud resources, the system must verify the identity of the visitor to make sure the visitor has the related permissions. To access the cloud resources, you must obtain the AccessKey ID and AccessKey Secret of your personal account to authorize your logon.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the upper-right corner, click your avatar and select Personal Information.
3. In the Alibaba Cloud AccessKey section, view the AccessKey information of your personal account.

> **Note:**
> AccessKey ID and AccessKey Secret are keys for you to access the cloud resources with full permissions. Keep them properly.

1.10.3 View third-party AccessKey

If a third-party application calls the cloud control platform, you must obtain the AccessKey ID and AccessKey Secret used to authorize the logon of the third-party application.

Procedure

1. *Log on to the Apsara Stack console*.

2. In the upper-right corner, click your avatar and select Personal Information.
3. In the Third-Party AccessKey section, view the third-party AccessKey information.

> **Note:**
> AccessKey ID and AccessKey Secret are keys for you to call the cloud control platform with full permissions. Keep them properly.

1.10.4 Change your avatar

You can change your avatar in the system by selecting a default avatar or uploading a custom avatar.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the upper-right corner of the page, click your avatar and select Personal Information.
3. On the Personal information Page, click Change Profile Picture under the avatar.
4. In the Change Profile Picture dialog box, change your avatar.
   - Click the Default Avatar tab. Click the avatar and then click OK. The avatar is changed.
   - Click the Custom Avatar tab. Click Upload Files. Select the picture and then click Open. The picture appears in the preview section. Crop the picture as required. Then, click OK. The avatar is changed.

1.10.5 Change your logon password

To improve security, you must change your logon password in time.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the upper-right corner, click your avatar and select Personal Information.
3. On the Personal Information page, click Change Logon Password under the avatar.
4. In the Change Logon Password dialog box, enter the current password, new password, and confirm password.
5. Then, click OK.

2 Elastic Compute Service (ECS)

2.1 What is ECS?

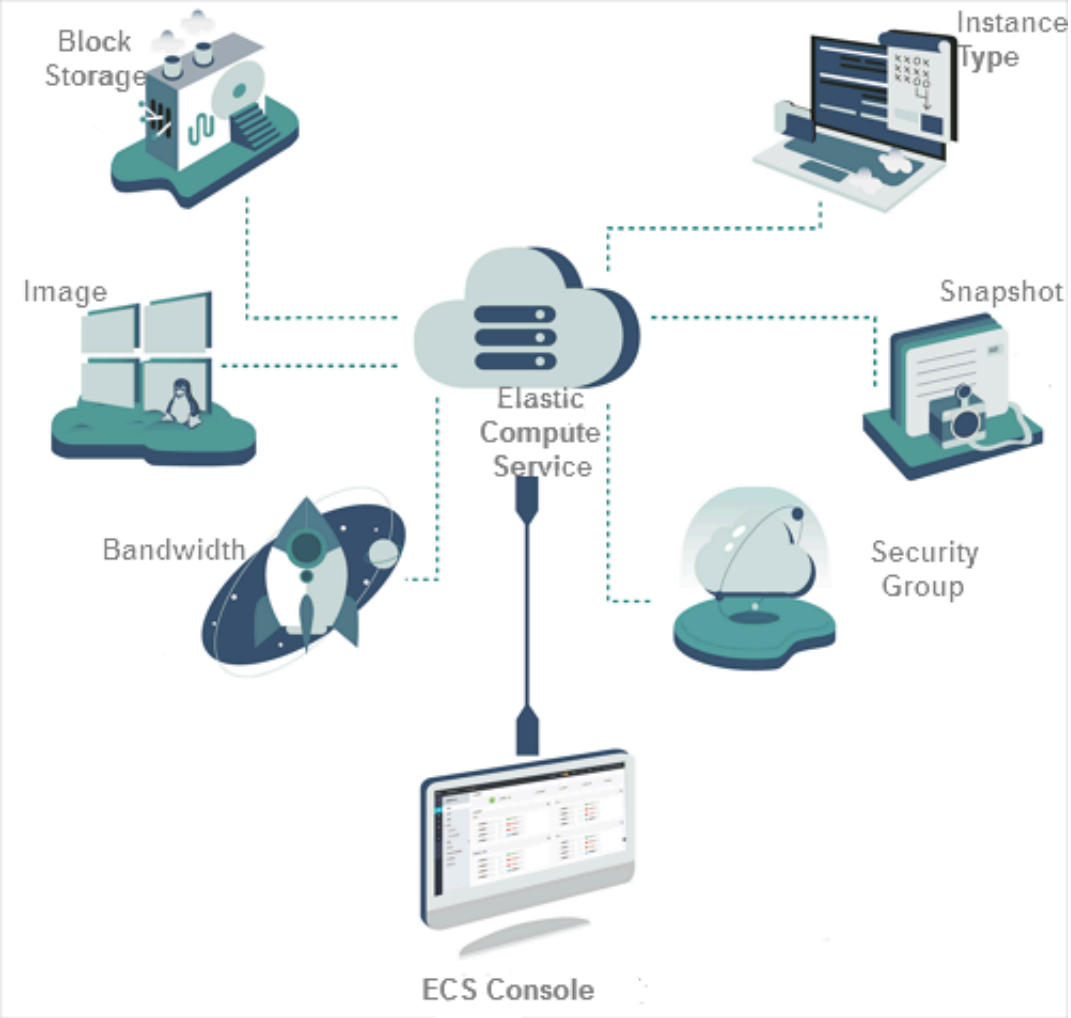
2.1.1 Overview

Elastic Compute Service (ECS) is a type of computing service that features elastic processingcapabilities. Compared with physical servers, ECS can be more efficientlymanaged and is more user-friendly. You can create instances, resize disks, and add or release any number of ECS instances at any time based on your business needs.

An ECS instance is a virtual computing environment that contains the most basic components of computers such as the CPU, memory, and storage. Users perform operations on ECS instances. Instances are core components of ECS, and operations can be performed on instances by using the ECS console. Other resources such as

block storage, images, and snapshots can only be used after they are integrated into ECS instances. For more information, see *Figure 2-1: ECS components*.

Figure 2-1: ECS components



2.1.2 Instance types

An instance is the smallest unit that can provide compute capabilities and services for your business. The compute capabilities vary with instance type.

The ECS instance type defines the basic properties of an ECS instance: CPU (including CPU model and clock speed) and memory. When you create an instance , you must also configure the block storage, image, and network type in addition to the instance type. The following table describes all instance families and their types.

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	Private IP addresses per ENI
n4	ecs.n4.small	1	2.0	N/A	0.5	50	1	1
	ecs.n4.large	2	4.0	N/A	0.5	100	1	1
	ecs.n4.xlarge	4	8.0	N/A	0.8	150	1	2
	ecs.n4.2xlarge	8	16.0	N/A	1.2	300	1	2
	ecs.n4.4xlarge	16	32.0	N/A	2.5	400	1	2
	ecs.n4.8xlarge	32	64.0	N/A	5.0	500	1	2
mn4	ecs.mn4.small	1	4.0	N/A	0.5	50	1	1
	ecs.mn4.large	2	8.0	N/A	0.5	100	1	1
	ecs.mn4.xlarge	4	16.0	N/A	0.8	150	1	2
	ecs.mn4.2xlarge	8	32.0	N/A	1.2	300	1	2
	ecs.mn4.4xlarge	16	64.0	N/A	2.5	400	1	2
	ecs.mn4.8xlarge	32	128.0	N/A	5.0	500	2	8
xn4	ecs.xn4.small	1	1.0	N/A	0.5	50	1	1
e4	ecs.e4.small	1	8.0	N/A	0.5	50	1	1
	ecs.e4.large	2	16.0	N/A	0.5	100	1	1
	ecs.e4.xlarge	4	32.0	N/A	0.8	150	1	2

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	Private IP addresses per ENI
	ecs.e4.2xlarge	8	64.0	N/A	1.2	300	1	3
	ecs.e4.4xlarge	16	128.0	N/A	2.5	400	1	8
sn1ne	ecs.sn1ne.large	2	4.0	N/A	1.0	300	2	2
	ecs.sn1ne.xlarge	4	8.0	N/A	1.5	500	2	3
	ecs.sn1ne.2xlarge	8	16.0	N/A	2.0	1,000	4	4
	ecs.sn1ne.3xlarge	12	24.0	N/A	2.5	1,300	4	6
	ecs.sn1ne.4xlarge	16	32.0	N/A	3.0	1,600	4	8
	ecs.sn1ne.6xlarge	24	48.0	N/A	4.5	2,000	6	8
	ecs.sn1ne.8xlarge	32	64.0	N/A	6.0	2,500	8	8
sn2ne	ecs.sn2ne.large	2	8.0	N/A	1.0	300	2	2
	ecs.sn2ne.xlarge	4	16.0	N/A	1.5	500	2	3
	ecs.sn2ne.2xlarge	8	32.0	N/A	2.0	1,000	4	4

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	Private IP addresses per ENI
	ecs.sn2ne.3xlarge	12	48.0	N/A	2.5	1,300	4	6
	ecs.sn2ne.4xlarge	16	64.0	N/A	3.0	1,600	4	8
	ecs.sn2ne.6xlarge	24	96.0	N/A	4.5	2,000	6	8
	ecs.sn2ne.8xlarge	32	128.0	N/A	6.0	2,500	8	8
	ecs.sn2ne.14xlarge	56	224.0	N/A	10.0	4,500	14	8
se1ne	ecs.se1ne.large	2	16.0	N/A	1.0	300	2	2
	ecs.se1ne.xlarge	4	32.0	N/A	1.5	500	2	3
	ecs.se1ne.2xlarge	8	64.0	N/A	2.0	1,000	4	4
	ecs.se1ne.3xlarge	12	96.0	N/A	2.5	1,300	4	6
	ecs.se1ne.4xlarge	16	128.0	N/A	3.0	1,600	4	8
	ecs.se1ne.6xlarge	24	192.0	N/A	4.5	2,000	6	8

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	Private IP addresses per ENI
	ecs.se1ne.8xlarge	32	256.0	N/A	6.0	2,500	8	8
	ecs.se1ne.14xlarge	56	480.0	N/A	10.0	4,500	14	8
se1	ecs.se1.large	2	16.0	N/A	0.5	100	1	2
	ecs.se1.xlarge	4	32.0	N/A	0.8	200	1	3
	ecs.se1.2xlarge	8	64.0	N/A	1.5	400	1	4
	ecs.se1.4xlarge	16	128.0	N/A	3.0	500	2	8
	ecs.se1.8xlarge	32	256.0	N/A	6.0	800	3	8
	ecs.se1.14xlarge	56	480.0	N/A	10.0	1,200	4	8
ebmg5	ecs.ebmg5.24xlarge	96	384.0	N/A	10.0	4,000	8	32
i2	ecs.i2.xlarge	4	32.0	1 × 894	1.0	500	2	3
	ecs.i2.2xlarge	8	64.0	1 × 1,788	2.0	1,000	2	4
	ecs.i2.4xlarge	16	128.0	2 × 1,788	3.0	1,500	4	8
	ecs.i2.8xlarge	32	256.0	4 × 1,788	6.0	2,000	8	8
	ecs.i2.16xlarge	64	512.0	8 × 1,788	10.0	4,000	16	8
d1	ecs.d1.2xlarge	8	32.0	4 × 5,500	3.0	300	1	4

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	Private IP addresses per ENI
	ecs.d1.3xlarge	12	48.0	6 × 5,500	4.0	400	1	6
	ecs.d1.4xlarge	16	64.0	8 × 5,500	6.0	600	2	8
	ecs.d1.6xlarge	24	96.0	12 × 5,500	8.0	800	2	8
	ecs.d1-c8d3.8xlarge	32	128.0	12 × 5,500	10.0	1,000	4	8
	ecs.d1.8xlarge	32	128.0	16 × 5,500	10.0	1,000	4	8
	ecs.d1-c14d3.14xlarge	56	160.0	12 × 5,500	17.0	1,800	6	8
	ecs.d1.14xlarge	56	224.0	28 × 5,500	17.0	1,800	6	8
ecs.d2-zyy	ecs.d2-zyy-d0.4xlarge	16	64.0	N/A	3.0	300	2	8
	ecs.d2-zyy-d0.6xlarge	24	96.0	N/A	4.0	400	2	8
	ecs.d2-zyy-m40.12xlarge	48	160.0	12 x 7,500	10.0	1,000	4	8
sccg5ib	ecs.sccg5ib.24xlarge	96	384.0	N/A	10.0	4,500	8	32
scch5ib	ecs.scch5ib.16xlarge	64	192.0	N/A	10.0	4,500	8	32
sn1	ecs.sn1.medium	2	4.0	N/A	0.5	100	1	2

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	Private IP addresses per ENI
	ecs.sn1.large	4	8.0	N/A	0.8	200	1	3
	ecs.sn1.xlarge	8	16.0	N/A	1.5	400	1	4
	ecs.sn1.3xlarge	16	32.0	N/A	3.0	500	2	8
	ecs.sn1.7xlarge	32	64.0	N/A	6.0	800	3	8
sn2	ecs.sn2.medium	2	8.0	N/A	0.5	100	1	2
	ecs.sn2.large	4	16.0	N/A	0.8	200	1	3
	ecs.sn2.xlarge	8	32.0	N/A	1.5	400	1	4
	ecs.sn2.3xlarge	16	64.0	N/A	3.0	500	2	8
	ecs.sn2.7xlarge	32	128.0	N/A	6.0	800	3	8
	ecs.sn2.14xlarge	56	224.0	N/A	10.0	1,200	4	8

Instanc family	Instanc type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	Private IP address per ENI	FPGAs
f1	ecs.f1-c8f1.2xlarge	8	60.0	N/A	3.0	400	4	4	Intel Arria 10 GX 1150

Instanc family	Instanc type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	Private IP address per ENI	FPGAs
	ecs.f1-c8f1.4xlarge	16	120.0	N/A	5.0	1,000	4	8	2 × Intel Arria 10 GX 1150
	ecs.f1-c28f1.7xlarge	28	112.0	N/A	5.0	2,000	8	8	Intel Arria 10 GX 1150
	ecs.f1-c28f1.14xlarge	56	224.0	N/A	10.0	2,000	14	8	2 × Intel Arria 10 GX 1150
f3	ecs.f3-c16f1.4xlarge	16	64.0	N/A	5.0	1,000	4	8	1 × Xilinx VU9P
	ecs.f3-c16f1.8xlarge	32	128.0	N/A	10.0	2,000	8	8	2 × Xilinx VU9P
	ecs.f3-c16f1.16xlarge	64	256.0	N/A	20.0	2,000	16	8	4 × Xilinx VU9P

Instanc family	Instanc type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	Private IP address per ENI	GPUs
gn5	ecs.gn5-c4g1.xlarge	4	30.0	440	3.0	300	1	3	1 × NVIDIA P100

Instanc family	Instanc type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	Private IP address per ENI	GPUs
	ecs.gn5-c8g1.2xlarge	8	60.0	440	3.0	400	1	4	1 × NVIDIA P100
	ecs.gn5-c4g1.2xlarge	8	60.0	880	5.0	1,000	2	4	2 × NVIDIA P100
	ecs.gn5-c8g1.4xlarge	16	120.0	880	5.0	1,000	4	8	2 × NVIDIA P100
	ecs.gn5-c28g1.7xlarge	28	112.0	440	5.0	1,000	8	8	1 × NVIDIA P100
	ecs.gn5-c8g1.8xlarge	32	240.0	1,760	10.0	2,000	8	8	4 × NVIDIA P100
	ecs.gn5-c28g1.14xlarge	56	224.0	880	10.0	2,000	14	8	2 × NVIDIA P100
	ecs.gn5-c8g1.14xlarge	54	480.0	3,520	25.0	4,000	14	8	8 × NVIDIA P100
gn4	ecs.gn4-c4g1.xlarge	4	30.0	N/A	3.0	300	1	3	1 × NVIDIA M40

Instanc family	Instanc type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	Private IP address per ENI	GPUs
	ecs.gn4-c8g1.2xlarge	8	30.0	N/A	3.0	400	1	4	1 × NVIDIA M40
	ecs.gn4.8xlarge	32	48.0	N/A	6.0	800	3	8	1 × NVIDIA M40
	ecs.gn4-c4g1.2xlarge	8	60.0	N/A	5.0	500	1	4	2 × NVIDIA M40
	ecs.gn4-c8g1.4xlarge	16	60.0	N/A	5.0	500	1	8	2 × NVIDIA M40
	ecs.gn4.14xlarge	56	96.0	N/A	10.0	1,200	4	8	2 × NVIDIA M40
ga1	ecs.ga1.xlarge	4	10.0	1 × 87	1.0	200	1	3	0.25 × AMD S7150
	ecs.ga1.2xlarge	8	20.0	1 × 175	1.5	300	1	4	0.5 × AMD S7150
	ecs.ga1.4xlarge	16	40.0	1 × 350	3.0	500	2	8	1 × AMD S7150
	ecs.ga1.8xlarge	32	80.0	1 × 700	6.0	800	3	8	2 × AMD S7150
	ecs.ga1.14xlarge	56	160.0	1 × 1,400	10.0	1,200	4	8	4 × AMD S7150

Instanc family	Instanc type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	Private IP address per ENI	GPUs
gn5i	ecs.gn5i-c2g1.large	2	8.0	N/A	1.0	100	2	2	1 × NVIDIA P4
	ecs.gn5i-c4g1.xlarge	4	16.0	N/A	1.5	200	2	3	1 × NVIDIA P4
	ecs.gn5i-c8g1.2xlarge	8	32.0	N/A	2.0	400	4	4	1 × NVIDIA P4
	ecs.gn5i-c16g1.4xlarge	16	64.0	N/A	3.0	800	4	8	1 × NVIDIA P4
	ecs.gn5i-c16g1.8xlarge	32	128.0	N/A	6.0	1,200	8	8	2 × NVIDIA P4
	ecs.gn5i-c24g1.12xlarge	48	192.0	N/A	10.0	2,000	8	8	2 × NVIDIA P4
	ecs.gn5i-c28g1.14xlarge	56	224.0	N/A	10.0	2,000	14	8	2 × NVIDIA P4
gn5e	ecs.gn5e-c11g1.3xlarge	10	58.0	N/A	2.0	150	1	6	1 × NVIDIA P4

Instanc family	Instanc type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	Private IP address per ENI	GPUs
	ecs. gn5e-c11g1. 5xlarge	22	116.0	N/A	4.0	300	1	8	2 × NVIDIA P4
	ecs. gn5e-c11g1. 11xlarge	44	232.0	N/A	6.0	600	2	8	4 × NVIDIA P4
	ecs. gn5e-c11g1. 22xlarge	88	464.0	N/A	10.0	1,200	4	15	8 × NVIDIA P4
gn6i	ecs. gn6i-c10g1. 2xlarge	10	42.0	N/A	5.0	800	2	4	1 × T4
	ecs. gn6i-c10g1. 5xlarge	20	84.0	N/A	8.0	1,000	4	6	2 × T4
	ecs. gn6i-c10g1. 10xlarge	40	168.0	N/A	15.0	2,000	8	8	4 × T4
	ecs. gn6i-c10g1. 20xlarge	80	336.0	N/A	30.0	4,000	16	8	8 × T4

The following instance types are only applicable to environments that are upgraded from Apsara Stack V2 to V3.

Instance family	Instance type	vCPUs	Memory (GiB)
n1	ecs.n1.tiny	1	1.0

Instance family	Instance type	vCPUs	Memory (GiB)
	ecs.n1.small	1	2.0
	ecs.n1.medium	2	4.0
	ecs.n1.large	4	8.0
	ecs.n1.xlarge	8	16.0
	ecs.n1.3xlarge	16	32.0
	ecs.n1.7xlarge	32	64.0
n2	ecs.n2.small	1	4.0
	ecs.n2.medium	2	8.0
	ecs.n2.large	4	16.0
	ecs.n2.xlarge	8	32.0
	ecs.n2.3xlarge	16	64.0
	ecs.n2.7xlarge	32	128.0
e3	ecs.e3.small	1	8.0
	ecs.e3.medium	2	16.0
	ecs.e3.large	4	32.0
	ecs.e3.xlarge	8	64.0
	ecs.e3.3xlarge	16	128.0
c1	ecs.c1.small	8	8.0
	ecs.c1.large	8	16.0
c2	ecs.c2.medium	16	16.0
	ecs.c2.large	16	32.0
	ecs.c2.xlarge	16	64.0
m1	ecs.m1.medium	4	16.0
	ecs.m1.xlarge	8	32.0
m2	ecs.m2.medium	4	32.0
s1	ecs.s1.small	1	2.0
	ecs.s1.medium	1	4.0
	ecs.s1.large	1	8.0
s2	ecs.s2.small	2	2.0
	ecs.s2.large	2	4.0

Instance family	Instance type	vCPUs	Memory (GiB)
	ecs.s2.xlarge	2	8.0
	ecs.s2.2xlarge	2	16.0
s3	ecs.s3.medium	4	4.0
	ecs.s3.large	4	8.0
t1	ecs.t1.small	1	1.0

2.1.3 Instance lifecycle

The lifecycle of an ECS instance begins when it is created and ends when it is released. This topic describes the instance status, status attributes, and corresponding API status.

An instance has several inherent states throughout its lifecycle, as shown in *Table 2-1: Lifecycle description*.

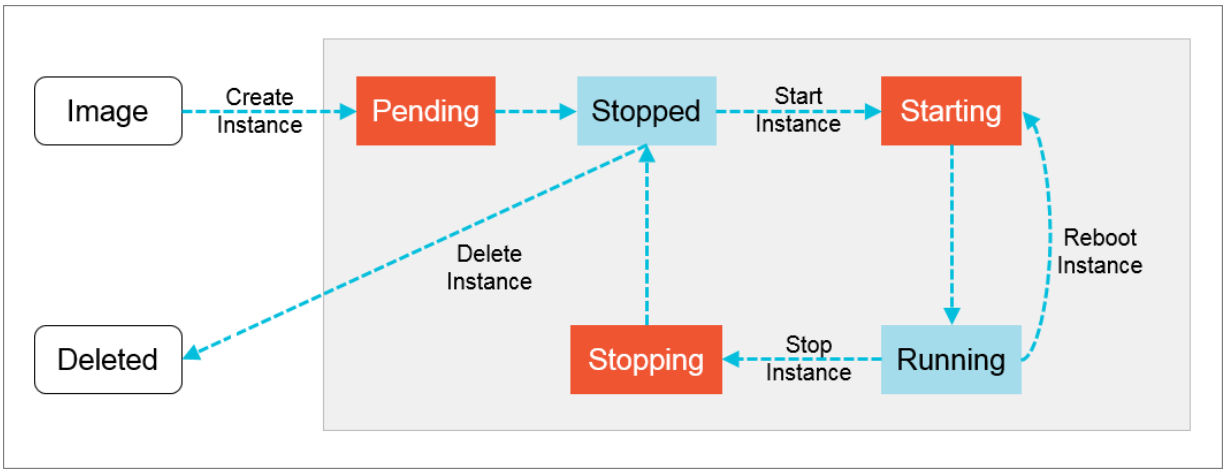
Table 2-1: Lifecycle description

Status	Status attribute	Description	Corresponding API status
`Instance being created`	**Intermediate**	**The instance is being created and is waiting to be enabled. If an instance remains in this status for a long period of time, an exception occurs.**	`Pending`
`Starting`	**Intermediate**	**After an instance is restarted or started from the console or through APIs, the instance enters the starting state before entering the `running` state. If an instance remains in the starting state for a long period of time, an exception occurs.**	`Starting`
`Running`	**Stable**	**Indicates that the instance is running normally and can accommodate your business needs.**	`Running`

Status	Status attribute	Description	Corresponding API status
Stopping	Intermediate	After an instance is stopped from the console or through APIs, the instance enters the stopping state before entering the stopped state. If an instance remains in the stopping state for a long period of time, an exception occurs.	Stopping
Stopped	Stable	Indicates that an instance has been stopped. An instance in the stopped state cannot provide external services.	Stopped
Reinitializing	Intermediate	After the system disk or data disk is reinitialized from the console or through APIs the instance enters the reinitializing state before entering the running state. If an instance remains in the reinitializing state for a long period of time, an exception occurs.	Stopped
Changing system disk	Intermediate	After the system disk is changed from the console or through APIs, the instance enters the changing system disk state before entering the running state. If an instance remains in the changing system disk state for a long time, an exception occurs.	Stopped

Table 2-1: Lifecycle description **describes corresponding relationship between instance states in the console and instance states in APIs.** *Figure 2-2: Instance status in APIs* **shows the instance states in APIs.**

Figure 2-2: Instance status in APIs



2.2 Instructions

2.2.1 Overview

Learn about precautions and limits before using ECS.

2.2.2 Restrictions

Learn about restrictions before performing operations on ECS instances.

- **Do not upgrade the kernel or operating system version of an ECS instance.**
- **Do not start SELinux for Linux systems except CentOS and RedHat.**
- **Do not detach PVDriver.**
- **Do not arbitrarily modify the MAC address of the network interface.**

2.2.3 Suggestions

Consider the following suggestions to make more efficient use of ECS:

- **ECS instances with 4 GiB or higher memory must use a 64-bit operating system. 32-bit operating systems have a maximum of 4 GiB of memory addressing.**
- **A 32-bit Windows operating system supports a maximum of 4 CPU cores.**
- **To ensure service continuity and avoid failover-induced service unavailability, we recommend that you configure service applications to boot automatically at system startup.**

2.2.4 Limits

Before using ECS instances, you must be familiar with the limits of instance type families.

General limits

- **Windows operating systems support a maximum of 64 vCPUs in instance specifications.**
- **Virtualization software installation and subsequent virtualization such as VMware are not supported.**
- **Currently, sound card applications are not supported. Only GPU instances support virtual sound cards. External hardware devices, such as hardware dongles, USB flash drives, external hard disks, and security tokens, cannot be directly connected to ECS instances.**
- **ECS does not support multicast protocols. If multicasting services are required, we recommend that you use unicast instead.**

Instance type family ga1

To create a ga1 instance, you must use the following images pre-installed with drivers:

- **Ubuntu16.04 with an AMD GPU driver pre-installed**
- **Windows Server 2016 English version with an AMD GPU driver pre-installed**
- **Windows 2008 R2 English version with an AMD GPU driver pre-installed**

Notes:

- **A ga1 instance uses an optimized driver provided by Alibaba Cloud and AMD . The driver is installed in images provided by Alibaba Cloud and is currently unavailable for download.**
- **If the GPU driver malfunctions due to improper removal of related components, you need to** *Change system disk* **to restore GPU related functions.**

  Note:
  This operation causes data loss.
- **If the driver malfunctions because an improper image is selected, you need to** *Change system disk* **to reselect an image with an AMD GPU driver pre-installed.**

- **For Windows 2008 or earlier versions, you cannot Connect to management terminal after the GPU driver takes effect. The management terminal is irresponsive with a black screen or stuck at the splash screen. You can access the system through other protocols, such as the remote desktop protocol (RDP) of Windows.**
- **RDP does not support DirectX, OpenGL and other related applications. You need to install the management terminal and a client, or use other supported protocols such as PCOIP and XenDesktop HDX 3D.**

Instance type family gn4

- **Bandwidth: select a bandwidth as needed.**

  Note:
  For Windows 2008 R2, you cannot Connect to management terminal after the GPU driver takes effect. You need to set the bandwidth to a non-zero value or attach an EIP to the created instance.

- **Image: select an image as needed.**

  If an NVIDIA GPU driver is not required, you can select any image, and *Install the CUDA and GPU drivers for a Linux instance* **or** *Install the CUDA and GPU drivers for a Windows instance***.**

Instance type family gn5i and gn5

- **Bandwidth: select a bandwidth as needed.**

  Note:
  For Windows 2008 R2, you cannot Connect to management terminal after the GPU driver takes effect. You need to set the bandwidth to a non-zero value or attach an EIP to the created instance.

- **Image: select an image as needed.**

  If an NVIDIA GPU driver is not required, you can select any image, and *Install the CUDA and GPU drivers for a Linux instance* **or** *Install the CUDA and GPU drivers for a Windows instance***.**

2.2.5 Notice for Windows users

Before using Windows-based ECS instances, you must consider the following points:

- **Data loss may occur if a local disk is used as the data disk of an instance. We recommend that you use a cloud disk to create your instance if you are not sure about the reliability of the data architecture.**
- **Do not close the built-in shutdownmon.exe process. Otherwise, the server may require a longer time to restart.**
- **Do not rename, delete, or disable Administrator accounts or it may affect the use of the server.**
- **We do not recommend that you use virtual memory.**
- **When you modify your computer name, you must synchronize the following key values in the registry. Otherwise, the computer name cannot be modified, causing failure when installing certain third-party programs. The following key values must be modified in the registry:**

```
HKEY_LOCAL_MACHINE\SYSTEM\ControlSet001\Control\ComputerName\
ActiveComputerName
```

```
HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Control\ComputerName\
ComputerName
```

2.2.6 Notice for Linux users

Before using Linux-based ECS instances, you must consider the following points:

- **Do not modify content of the default /etc/issue files under a Linux instance. Otherwise, the custom image created from the instance cannot be recognized, and instances created based on the image cannot start as expected.**
- **Do not arbitrarily modify the permissions of each directory in the partition where the root directory is located, especially permissions of /etc, /sbin, /bin, /boot, /dev, /usr, and /lib directories. Improper modification of permissions can cause errors.**
- **Do not rename, delete, or disable Linux root accounts.**
- **Do not compile or perform any other operations on the Linux kernel.**
- **We do not recommend the use of Swap for partitioning.**
- **Do not enable the NetWorkManager service. This service conflicts with the internal network service of the system, causing network errors.**

2.2.7 Notice on defense against DDoS attacks

You need to purchase Anti-DDoS Pro to defend against DDoS attacks. For more information, see *Apsara Stack Security Product Introduction*.

2.3 Quick start

2.3.1 Overview

This topic is designed to guide you through the preparation of instances. It describes how to quickly log on to the ECS console, create a security group, create an instance, and more.

You need to perform the following steps before using ECS:

1. *Create a security group*

   A security group is a virtual firewall used to control traffic to and from ECS instances. Each ECS instance must be added to at least one security group. Before creating an instance, you need to select a security group to control traffic to and from the instance. If no default security group exists, you must create one.

2. *Create an instance*

   An ECS instance is a virtual computer that consists of basic components such as CPU, memory, operating system, network, and disk. After creating a security group, you can select different *Instance types***to create instances based on your requirements.**

3. *Connect to an instance*

   You can connect to instance and install applications based on the network configuration and operating system of the ECS instance and your local operating system.

2.3.2 Log on to the ECS console

This topic describes how to log on to the ECS console.

Prerequisites

- **Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://**`IP address or domain name of the Apsara Stack console`**/manage.**

- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click LOGIN to go to the Dashboard page.
5. Log on to Apsara Stack console. Choose  > Compute, Storage & Networking > Elastic Compute Service from the top navigation bar.

2.3.3 Create a security group

Before creating an ECS instance in a VPC, you must first create a security group. A security group controls access to ECS instances.

Procedure

1. *Log on to the ECS console*.
2. On the Security Groups page, click Create Security Group.

3. **On the Create Security Group page, configure parameters of the security group.**

   Table 2-2: Security group parameters **describes the parameters.**

   Table 2-2: Security group parameters

Area	Parameter	Description
Region	Region	Required. The region to which the security group belongs. It must be the same region as the VPC.
Basic Settings	Department	Required. The department to which the security group belongs. It must be the same department as the VPC.
	Project	Optional. The project to which the security group belongs.
	Network Type	Required. The default network type is VPC. It must be the VPC to which the security group belongs.
	Security Group Name	Required. The name of the security group.
	Description	Optional. The description of the security group.

4. **Click OK.**

2.3.4 Create an instance

After creating a security group, you need to create an instance.

Prerequisites

- **Before creating an instance, you must first create a VPC and a VSwitch. For more information, see** *VPC User Guide*.
- **Ensure that at least one security group is available. If you do not have any security groups, see** *Create a security group*.
- **Before you create a GPU instance, see** *Limits*.

Procedure

1. *Log on to the ECS console*.
2. **On the Instances page, click Create Instance.**

3. **On the Create Instance page, configure the parameters of the instance.**

   Table 2-3: Instance parameters **describes the instance parameters.**

   Table 2-3: Instance parameters

Category	Parameter	Description
Region	Region	The region where the ECS instance resides.
	Zone	A zone is a physical area with separate power circuits and networks within a region. Zones are interconnected over the intranet, and fault isolation is implemented between different zones. If you need to increase the availability of your applications, we recommend that you create multiple instances in different zones.
Basic Settings	Department	The department to which the ECS instance belongs.
	Project	The project to which the ECS instance belongs.
Network	Network Type	Required. This default network type is VPC. Select a VPC name and a VSwitch name.
	Security Groups	Required. **Note:** Before creating an ECS instance, you must create a security group.
	Configure Private IP Address	Optional. The IP address must be within the CIDR block where the VSwitch is located. **Note:** · If you do not specify a private IP address, the system automatically allocates a private IP address for the instance. · If you specify a private IP address, you cannot create multiple instances at the same time.
Instance	Instance Series	The default value is Series 3.
	I/O Optimized	Optional. The parameter is set to I/O-Optimized Instance by default.

Category	Parameter	Description
	Select Instance Type	Optional. Select CPU and memory based on application requirements. Windows-based images require specific CPU and memory combinations. For more information, see *Suggestions*.
Image	Image Type	Required. Select `Public Image` or `Custom Image` as the image type of your system. Then, select the image to use.
Storage	System Disk	Required. Select an SSD or ultra cloud disk to install the operating system.
	Data Disk	Optional. You can select an SSD or ultra cloud disk. You can add a maximum of 16 data disks. The maximum capacity of each data disk is 32 TB. Select Release with Instance and Encrypt as needed. **Note:** • If `Release with Instance` is selected, the disk will be released when the instance is released, and the data cannot be recovered. If this parameter is not selected, the system will unmount the disk from the instance when the instance is released, but the data will be retained. • The created disk will be encrypted only if you select `Encrypt`. • If you prefer to add data disks later, you can follow the procedure described in *Create a disk*.
Password	Set Password	Optional. This parameter is set to Now by default. You can also select Later and set the password later in the ECS console by using the password reset function.
	Logon Password	If Now is selected, you set the logon password. The password must be 8 to 30 characters in length. The password must contain three of the following character types: uppercase letters, lowercase letters, digits, and special characters.

Category	Parameter	Description
	Confirm Password	Re-enter the logon password.  Note: The password is used to log on to the instance, not the management terminal.
Deployment Sets	Deployment Sets	The deployment set to which you want to add the instance.
Instance Name	Instance Name	Optional. We recommend that you set a meaningful name for your instance.  Note: The name must be 2 to 114 characters in length and can contain letters, digits, underscores (_), and hyphens (-). It must start with a letter.
Custom Data	Custom Data	You can enter the corresponding custom data encoding schemes. If the data to be entered is Base64 encoded, select Enter Base64 Encoded Information.  Note: Both bat and powershell are supported in Windows. When you use Base64 to encode custom data, make sure that [bat] or [powershell] appears as the first row. shell scripts are supported in Linux.
Instance Count	Instances	The default value is 1. If the value of this parameter is greater than 1, the instances are created in batches based on the configured parameters.  Note: A maximum of 50 ECS instances can be created at a time. You cannot create ECS instances in batches if a private IP address is specified in the instance configurations.

4. Click Create.

Result

Refresh the instance list to view the created instances. Instances in the `Running` state are created.

2.3.5 Connect to an instance

2.3.5.1 Overview

After you create an ECS instance, you can connect to the instance and install application software.

> **Note:**
> **The username of a Windows-based instance is Administrator, while that of a Linux-based instance is root.**

You can use either of the following methods to connect to and manage your ECS instance:

- **Use a remote connection tool to connect to the ECS instance.**

  > **Note:**
  > **This method only applies to instances that have EIPs.**

- **Click Connect to Management Terminal in the ECS console to connect to the ECS instance.**

2.3.5.2 Connect to a Linux-based instance by using SSH commands in Linux or Mac OS X

This topic describes how to use SSH commands to connect to a Linux-based instance.

Prerequisites

Create a security group and an instance.

Procedure

1. **Enter the following command: `ssh root@instance IP`.**
2. **Enter the password for the *root* user to log on to the instance.**

2.3.5.3 Connect to a Linux-based instance by using remote connection tools in Windows

This topic describes how to connect to an instance by using the PuTTY tool.

Prerequisites

Remote connection tools are designed with similar logics. In this example, PuTTY is used to connect to an instance. Download PuTTY at the following URL: *http://www.chiark.greenend.org.uk/~sgtatham/putty/*.

Procedure

1. Download and install PuTTY for Windows.
2. Start the PuTTY client and complete the following settings:
   - `Host Name (or IP Address)`: Enter the EIP of the instance to be connected.
   - `Port`: Select the default port 22.
   - `Connection Type`: Select `SSH`.
   - `Saved Session`: Enter the name of the session. Click Save. After the settings are saved, PuTTY remembers the name and IP address of the instance. This eliminates the need to enter them every time you connect to the instance.
3. Click Open to connect to the instance.

   When you connect to the instance for the first time, PuTTY displays security alerts. Click Yes to proceed.
4. Enter the username `root` and press Enter.
5. Enter the password for the instance and press Enter.

   If a message similar to the following one appears, a connection to the instance is established.

```
Welcome to aliyun Elastic Compute Server!
```

2.3.5.4 Connect to a Windows-based instance by using RDP

This topic describes how to connect to a Windows-based instance by using Remote Desktop Protocol (RDP).

Prerequisites

Create an instance and a security group, and ensure that the instance operating system is installed with CredSSP-related security updates.

Procedure

1. Activate Remote Desktop Connection through any of the following methods:
   - Click Start, enter Remote Desktop Connection in the search box, and click Remote Desktop Connection in the search result.
   - Enter `mstsc` in the search box and click mstsc in the search result.
   - Press `Windows Key+R`. In the Run dialog box that appears, enter `mstsc` and press Enter to activate Remote Desktop Connection.
2. In the Remote Desktop Connection dialog box, enter the EIP address of the instance and click Show Options (O).
3. Enter the username, which is `Administrator` by default. If you do not want to enter the password upon subsequent logons, select Allow me to save credentials (R). After completing the settings, click Connect to connect to the instance. You can also complete other settings as follows before you connect to the instance.

   You can also complete other settings as follows before you connect to the instance:
   - If you want to copy local texts to the instance, click the Local Resources tab and select Clipboard.
   - If you want to copy a local file to the instance, click the Local Resources tab. Click Details. Select Drivers and then the driver letter of the data disk where the file is to be stored. After completing the settings, click OK.
   - You can click the Display tab to adjust the size of the remote desktop. Typically, you can use the full screen mode.
4. In the dialog box that appears, enter the password for the `Administrator` account of the Windows-based instance and click OK to connect to the instance.

Result

If the Remote Desktop Connection window displays the Windows desktop, a connection to the instance is established.

> **Note:**
> If authentication errors occur or the required function is not supported, you need to install security updates before proceeding. The procedure is as follows: *Connect to the instance*. Choose Control Panel > System and Security > Windows Updates. Click Check Updates to view and install available updates. Restart the instance for the updates to take effect.

2.3.5.5 Connect to an ECS instance by clicking Connect to Management Terminal in the ECS console

When remote connection tools such as PuTTY, Xshell, and SecureCRT cannot be used, you can click Connect to Management Terminal in the ECS console to connect to an ECS instance.

Prerequisites

- **Create an instance and a security group.**
- **To use the management terminal, you must import the root certificate to the web browser. For more information, see** *Install a certificate*.
- **Before connecting to the management terminal, you must perform the following operations described in** *Change the management terminal password*.

> **Note:**
> **The management terminal password is used to log on to the management terminal of the ECS console, while the instance password is used to log on to the instance.**

Context

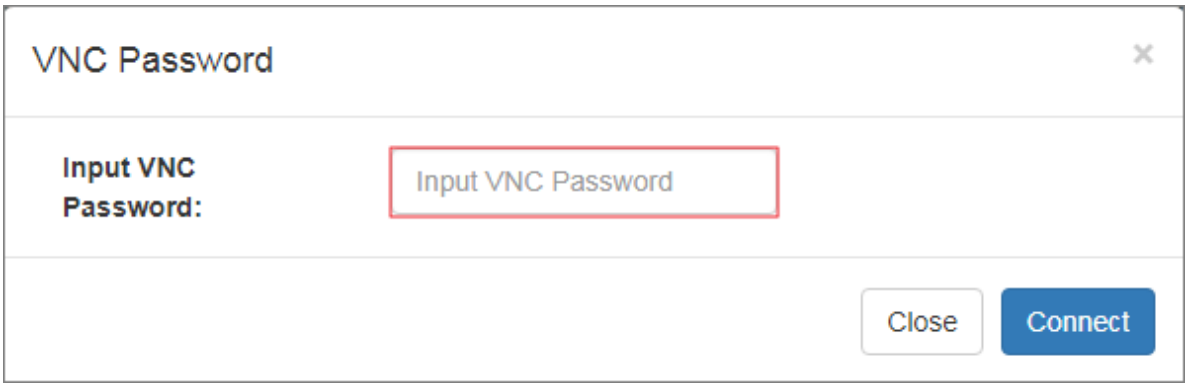
Connect to Management Terminal applies to the following scenarios (including but not limited to):

- **If it takes a long time to boot an instance (for example, self-test is initiated), you can view the progress by clicking Connect to Management Terminal.**
- **If a software-based remote connection fails due to incorrect instance configurations (for example, the firewall has been enabled by incorrect operations), you can connect to the instance by clicking Connect to Management Terminal and disable the firewall.**
- **If a remote connection fails due to high CPU or bandwidth usage (for example, botnet attacks occur), you can connect to the instance by clicking Connect to Management Terminal and terminate abnormal processes.**

Procedure

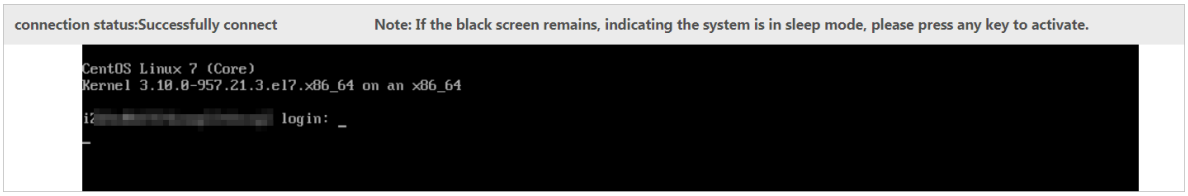
1. *Log on to the ECS console*.
2. **On the Instances page, click the** **icon in the** `Actions` **column corresponding to an instance, and choose View Details from the shortcut menu.**
3. **On the Instance Details page, click Connect to Management Terminal.**

4. In the dialog box that appears, enter the management terminal password and click Connect.



5. The logon page appears after you connect to the management terminal.

   The following figure shows the logon page in Linux.



6. Enter the username and password to log on to the instance.

   - For Linux-based instances, enter the `root` username and logon password.
   - For Windows-based instances, enter the `administrator` username and logon password.

> **Note:**
> There are no visual indicators when you enter the logon password in Linux. Press Enter after you enter the password.

2.4 Instances

2.4.1 Overview

An instance is the smallest unit in ECS that provides computing services. Computing capabilities vary with instance type.

2.4.2 View an instance

In the ECS console, you can view the existing instances and their details.

Prerequisites

For more information about how to create an instance, see *Create an instance*.

Procedure

1. *Log on to the ECS console*.
2. On the Instances page, set `Department`, `Region`, or enter `Instance Name`. Click Search to find the target instance.

   Note:

   - In the upper-right corner of the Instances page, click Filter Columns. In the Filter Columns dialog box that appears, select the desired attributes, and click OK. The selected attributes are displayed in the instance list.
   - On the Instances page, select other filtering conditions from the Instance Name drop-down list, including Instance ID, Instance State, VPC ID, IP Address, Project Name, Project ID, CPU Cores, Memory Size, and Image ID.
3. Click the icon in the `Actions` column corresponding to an instance, and choose View Details from the shortcut menu. On the Instance Details page that appears, you can view details of the instance.

   Note:
   On the Instances page, you can click an instance ID to go to the Instance Details page.

2.4.3 Edit an instance

You can modify the name and description of an existing instance in the ECS console.

Prerequisites

For more information about how to create an instance, see *Create an instance*.

Procedure

1. *Log on to the ECS console*.
2. Click the icon in the `Actions` column corresponding to an instance, and choose Edit from the shortcut menu.
3. In the dialog box that appears, set `Name`, `Description`, and `Custom Data`. Then, click OK.

   Note:

You can run Bat or PowerShell scripts for Windows-based ECS instances. Before being Base64 encoded, the scripts must use [bat] or [powershell] as the first line. You can run Shell scripts for Linux-based ECS instances.

2.4.4 Stop, restart, or start an instance

In the ECS console, you can stop, restart, or start an instance like operating a real server.

Prerequisites

For more information about how to create an instance, see *Create an instance*.

Procedure

1. *Log on to the ECS console*.
2. On the Instances page, click the icon in the `Actions` column corresponding to an instance, and choose Restart, Stop, or Start from the shortcut menu.
   - You can stop or restart an instance only when it is in the `Running` state, and start an instance only when it is in the `Stopped` state.

     Notice:
     The stop and restart operations stop your instance and interrupt your business. Exercise caution when performing the operations.
   - Click the icon in the Actions column corresponding to an instance and choose View Details from the shortcut menu. On the Instance Details page, click Restart, Stop or Start in the upper-right corner.

2.4.5 Delete an instance

You can delete unwanted instances in the console.

Prerequisites

The instances to be deleted must be in the `Stopped` state.

Procedure

1. *Log on to the ECS console*.

2. In the `Actions` column corresponding to an instance, click the icon and choose Delete from the shortcut menu. In the message that appears, click OK.

> **Note:**
> Alternatively, in the `Actions` column corresponding to an instance, click the icon and choose View Details from the shortcut menu. On the Instance Details page, click Delete. In the message that appears, click OK.

2.4.6 Change configurations

You can modify instance configurations in the ECS console.

Prerequisites

The instances to be modified must be in the `Stopped` state.

Procedure

1. *Log on to the ECS console*.
2. In the `Actions` column corresponding to an instance, click the icon and choose Change Configuration from the shortcut menu. Select new CPU and memory specifications and click OK.

> **Note:**
> *Restart the instance* in the console for the new configurations to take effect.

2.4.7 Change ownership

In the ECS console, you can change the department and project to which an instance belongs.

Procedure

1. *Log on to the ECS console*.
2. Click the icon in the `Actions` column corresponding to an instance, and choose Change Ownership from the shortcut menu.
3. In the Change Ownership dialog box that appears, reset `Department`, `Project`, or `Security Groups`. Then, click OK.

2.4.8 Change the password used to log on to the ECS instance

In the ECS console, you can change the instance logon password.

Context

If you forget to set the instance logon password upon instance creation, you can reset it through Change Password in the ECS console.

Procedure

1. *Log on to the ECS console*.
2. **Click the**  **icon in the Actions column corresponding to an instance, and choose View Details from the shortcut menu.**
3. **On the Instance Details page, click Change Password.**
4. **In the dialog box that appears, set** `Logon Password` **and** `Confirm Password` **to a new password. Then, click Submit.**
5. *Restart the instance* **in the console for the new password to take effect.**

2.4.9 Change the management terminal password

In the ECS console, you can change the management terminal password.

Procedure

1. *Log on to the ECS console*.
2. **Click the**  **icon in the Actions column corresponding to an instance, and choose View Details from the shortcut menu.**
3. **On the Instance Details page, click Change Management Terminal Password.**
4. **In the dialog box that appears, set** `Logon Password` **and** `Confirm Password` **to a new password. Then, click Submit.**

> **Notice:**
> *Restart the instance* **in the console for the new password to take effect.**

2.4.10 Add an ECS instance to a security group

In the ECS console, you can use two methods to add an instance to a security group.

Context

As an important means of isolation for network security, a security group functions as a virtual firewall that controls network access to ECS instances. In the ECS

console, you can add an instance to a maximum of five security groups. After you add an instance to a security group, the security group rules apply to the instance automatically.

Method 1: Add an instance to a security group from the Instances page

1. *Log on to the ECS console*.
2. **On the Instances page, click the icon in the `Actions` column corresponding to an instance, and choose View Details from the shortcut menu.**
3. **Click the Instance Security Group tab. On the displayed tab, click Join Security Group.**
4. **In the Add ECS Instance to Security Group dialog box that appears, select the target security group and click OK.**

Method 2: Add an instance to a security group from the Security Groups page

1. *Log on to the ECS console*.
2. **Click the Security Groups tab.**
3. **Click a security group ID. The ECS Instances page appears.**

> **Note:**
> **Alternatively, in the `Actions` column corresponding to a security group, click the icon, and choose View Details from the shortcut menu. The ECS Instances page appears.**

4. **In the upper-right corner of the ECS instances page, click Add ECS Instance.**
5. **In the Add ECS Instance to Security Group dialog box that appears, select an instance and click OK.**

2.4.11 Customize instance data

ECS allows you to run the instance customization script upon startup and import data into instances.

Context

The instance data customization feature is applicable to both Windows and Linux instances. It allows you to:

- **Run the instance customization script upon startup.**
- **Import data into instances.**

Usage instructions

- **Limits**

  The instance data customization feature can only be used when an instance meets all the following requirements:

  - **Network type: VPC**
  - **Image: a system image or a custom image that is inherited from the system image**
  - **Operating system: one type included in** *Table 2-4: Supported operating systems*

    Table 2-4: Supported operating systems

Windows	Linux
■ Windows Server 2016 64-bit ■ Windows Server 2012 64-bit ■ Windows Server 2008 64-bit	■ CentOS ■ Ubuntu ■ SUSE Linux Enterprise ■ OpenSUSE ■ Debian ■ Aliyun Linux

  - **When you configure instance data customization scripts, you must enter custom data based on the type of operating system and script.**

    Note:
    Only English characters are allowed.

    ■ **If your data is Base64 encoded, select Enter Base64 Encoded Information.**

    Note:
    The size of the customization script cannot exceed 16 KB before the data is Base64 encoded.

    ■ **For Linux instances, the script format must meet the requirements described in** *Types of Linux instance customization scripts*.

    ■ **For Windows instances, the script can only start with** `[bat]` **or** `[powershell]`.

- **After starting an instance, run a command to view the following information:**
  - **Execution result of the instance customization script**
  - **Data imported to instances**
- **Console: You can modify the custom instance data in the console. Whether the modified instance customization script needs to be re-executed depends on the script type. For example, if the `bootcmd` script in `Cloud Config` is modified for Linux instances, the script is automatically executed each time instances are restarted.**
- **OpenAPI: You can also use OpenAPI to customize instance data. For more information, see CreateInstance and ModifyInstanceAttribute in** *ECS Developer Guide* **.**

Linux instance data customization scripts

Linux instance data customization scripts provided by Alibaba Cloud are designed based on the cloud-init architecture. They are used to automatically configure parameters of Linux instances. Customization script types are compatible with the cloud-init.

Description of Linux instance data customization scripts

- **Linux instance customization scripts are executed after instances are started and before `/etc/init` is executed.**
- **Linux instance customization scripts can only be executed with root permissions by default.**

Types of Linux instance customization scripts

- **User-Data Script**
  - **Description: A script, such as shell script, is used to customize data.**
  - **Format: The first line must include `#!`, such as `#! /bin/sh`.**
  - **Limit: The script size (including the first line) cannot exceed 16 KB before the data is Base64 encoded.**
  - **Frequency: The script is executed only when instances are started for the first time.**
  - **Example:**

```
#! /bin/sh
```

```
echo "Hello World. The time is now $(date -R)!" | tee /root/
output10.txt
```

- **Cloud Config Data**
  - **Description: Predefined data is used to configure services, such as specifying yum sources or importing SSH keys.**
  - **Format: The first line must be `#cloud-config`.**
  - **Limit: The script size (including the first line) cannot exceed 16 KB before the data is Base64 encoded.**
  - **Frequency: The script execution frequency varies with the specific service.**
  - **Example:**

```
#cloud-config
apt:
primary:
- arches: [default]
uri: http://us.archive.ubuntu.com/ubuntu/
```

- **Include**
  - **Description: The configuration content can be saved in a text file and imported into cloud-init as a URL.**
  - **Format: The first line must be `#include`.**
  - **Limit: The script size (including the first line) cannot exceed 16 KB before the data is Base64 encoded.**
  - **Frequency: The script execution frequency depends on the script type in the text file.**
  - **Example:**

```
#include
```

```
http://ecs-image-test.oss-cn-hangzhou.aliyuncs.com/userdata/
cloudconfig
```

- **GZIP format**
  - **Description: Cloud-ini limits the size of customization scripts to 16 KB. You can compress and import the script file into the customization script if the file size exceeds 16 KB.**
  - **Format: The .gz file is imported into the customization script as a URL in `#include`.**
  - **Frequency: The script execution frequency depends on the script content contained in the GZIP file.**
  - **Example:**

```
#include
http://ecs-image-test.oss-cn-hangzhou.aliyuncs.com/userdata/config
.gz
```

View the custom data of a Linux instance

Run the following command in the instance:

```
curl http://100.100.100.200/latest/user-data
```

Windows instance customization scripts

Windows instance customization scripts independently developed by Alibaba Cloud can be used to initialize Windows instances.

There are two types of Windows instance customization scripts:

- **Batch processing program: starts with `[bat]` and serves as the first line. The script size must be smaller than 16 KB before the data is Base64 encoded.**
- **PowerShell script: starts with `[powershell]` and serves as the first line. The script size must be smaller than 16 KB before the data is Base64 encoded.**

View the custom data of a Windows instance

Run the following PowerShell command in the instance:

```
Invoke-RestMethod http://100.100.100.200/latest/user-data/
```

2.4.12 Modify private IP addresses

In the ECS console, you can modify the private IP address of an ECS instance.

Prerequisites

Before you modify the private IP address of an instance, ensure that the instance is in the stopped state. For more information about how to stop an instance, see *Stop an instance*.

Context

Each instance has a private NIC that is bound with a private IP address. The private IP address is based on the VSwitch CIDR block.

Procedure

1. *Log on to the ECS console*.
2. **Click the Instances tab. Click the**  **icon in the** `Actions` **column corresponding to the instance and choose Change Private IP Address from the shortcut menu.**
3. **In the Change Private IP Address dialog box that appears, enter a new** `Private IP address` **and click OK.**

2.4.13 Install a certificate

Before you log on to the Management Terminal, you must export the certificate from the site and install it in your local web browser.

Context

The Management Terminal feature is provided by the back-end VNC proxy service . The VNC proxy service uses a different certificate from that of Apsara Infrastructure Management Framework. You must manually import the certificate.

Procedure

1. **Export the certificate from the site.**
   a) **Log on to Apsara Stack console. Press F12 or Fn+F12 to view and select a certificate. For example, in the Chrome browser, press F12 to open the developer tools.**
   b) **In the Certificate dialog box, click Copy to File. Enter a name and save the certificate to your local machine.**
   c) **Click Finish. A message appears, indicating that the certificate is exported.**

2. Install the certificate in your local web browser.
   a) Double-click the certificate saved in your local machine. In the dialog box that appears, click Install Certificate.
   b) In the dialog box that appears, click Put all certificate in the following store and Browse. In the dialog box that appears, select Trusted Root Certificate Authority and click OK.
   c) After importing the certificate, click Finish.
3. Restart your web browser and log on to Apsara Stack console. A secure indicator in green is displayed in the left part of the address bar, indicating that the certificate is installed.

2.4.14 Install the CUDA and GPU drivers for a Linux instance

The device where a GPU instance runs must be installed with the GPU driver. If the image you use does not contain a pre-installed GPU driver, you must install the CUDA and GPU drivers for the instance.

Context

When installing NVIDIA drivers, you must install the kernel package that contains the kernel header file before you install the CUDA and GPU drivers.

Procedure

1. Install the kernel package.
   a) Run the `uname -r` command to view the current kernel version.

      A similar output is displayed:

      - CentOS: `3.10.0-862.14.4.el7.x86_64`
      - Ubuntu: `4.4.0-117-generic`

   b) Copy the kernel package of the corresponding version to the instance and install the package.

      - CentOS: Copy the RPM package of the `kernel-devel` component and run the `rpm -ivh 3.10.0-862.14.4.el7.x86_64.rpm` command to install the package. *`3.10.0-862.14.4.el7.x86_64.rpm`* is used as an example. Replace it with the actual package name.
      - Ubuntu: Copy the DEB package of the `linux-headers` component and run the `dpkg -i 4.4.0-117-generic.deb` command to install the package. *4*

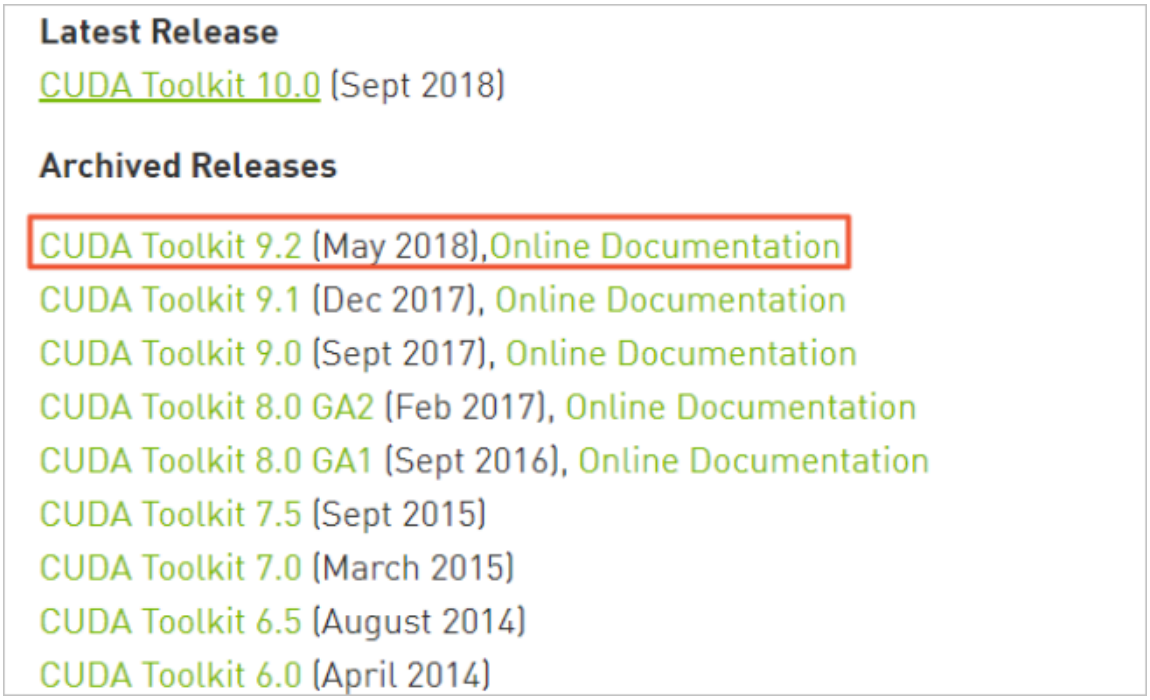
.4.0-117-generic.deb **is used as an example. Replace it with the actual package name.**

2. **Download the CUDA Toolkit.**

   a) **Access the *official download page*. Choose a suitable version based on the GPU application requirements for CUDA.**

   You can choose *CUDA Toolkit 9.2*.

   Figure 2-3: Download the CUDA Toolkit



   b) **Choose a platform based on your operating system. Set Installer Type to runfile (local) and click Download.**

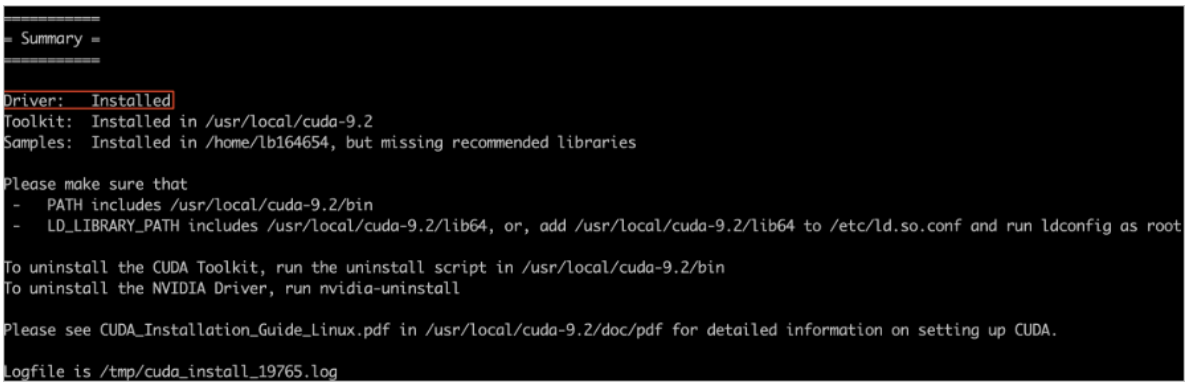
NVIDIA drivers are already included in the CUDA Toolkit.

Figure 2-4: Download drivers

3. **Copy the downloaded** *cuda_9.2.148_396.37_linux.run* **file to the instance.** *cuda_9.2.148_396.37_linux.run* **is used as an example. Replace it with the actual file name.**

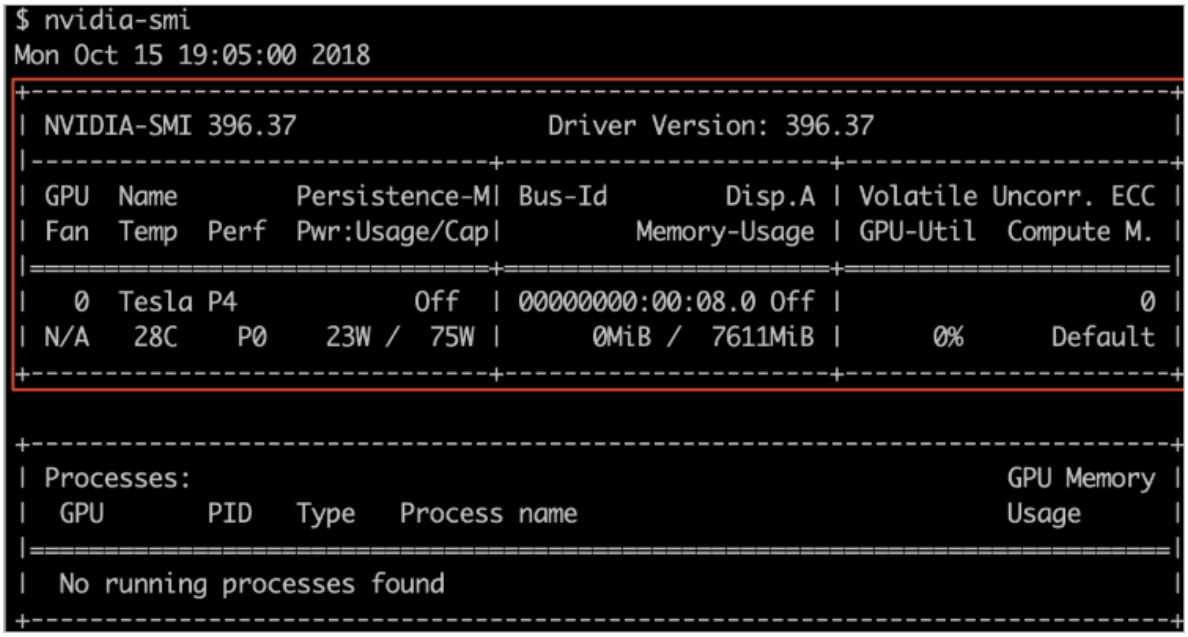
4. **Run the `sudo sh ./cuda_9.2.148_396.37_linux.run --silent --verbose --driver --toolkit --samples` command to install the CUDA. *`cuda_9.2.148_396.37_linux.run`* is used as an example. Replace it with the actual file name.**
   The installation takes 10 to 20 minutes. When `Driver: Installed` is displayed, the installation is successful.

   Figure 2-5: CUDA installation result



5. **Run the `nvidia-smi` command to view the GPU driver status.**
   If GPU driver details are displayed, the driver is in the normal state.

   Figure 2-6: View the GPU driver status



What's next

If you want to run the OpenGL program, you must first purchase the licenses and GRID drivers. For more information about the installation procedure, see official NVIDIA documentation.

2.4.15 Install the CUDA and GPU drivers for a Windows instance

The device where a GPU instance runs must be installed with the GPU driver. If the image you use does not contain a pre-installed GPU driver, you must install the CUDA and GPU drivers for the instance.

Context

If you want to compile CUDA programs, first install a Windows compiling environment, such as Visual Studio 2015. If you do not need to compile CUDA programs, ignore it.

Procedure

1. **Download the CUDA Toolkit.**
   a) **Access the *official download page*. Choose a suitable version based on the GPU application requirements for CUDA.**
      You can choose *CUDA Toolkit 9.2*.
   b) **Choose a platform based on your operating system. Set Installer Type to exe (local) and click Download.**
      NVIDIA drivers are already included in the CUDA Toolkit.
2. **Copy the downloaded *Cuda_9.2.148_windows.exe* file to the instance. *cuda_9.2.148_windows.exe* is used as an example. Replace it with the actual name of the downloaded file.**
3. **Double-click *cuda_9.2.148_windows.exe* and follow the installation wizard to install the CUDA. *cuda_9.2.148_windows.exe* is used as an example. Replace it with the actual name of the downloaded file.**
   The installation takes 10 to 20 minutes. When `Installed: - Nsight Monitor and HUD Launcher` is displayed, the installation is successful.
4. **Press `Win+R` and enter `devmgmt.msc`.**
   The NVIDIA device is displayed in Display Adapter.
5. **Press `Win+R` and enter `cmd`. Run the `C:\Program Files\NVIDIA Corporation\NVSMI\nvidia-smi` command.**
   If GPU driver details are displayed, the driver is in the normal state.

What's next

If you want to run OpenGL and DirectX programs, you must first purchase the licenses and GRID drivers. For more information about the installation procedure, see official NVIDIA documentation.

2.5 Disks

2.5.1 Overview

For ECS instances, a cloud disk can be seen as a physical disk. You have to attach and format a cloud disk before using it.

Disk types

ECS disks are classified into basic disks, SSD disks, and ultra disks. A mount point refers to the position of an ECS disk on the disk controller bus. The selected mount point corresponds to the disk device number in Linux, and is consistent with the disk sequence in the disk manager in Windows.

Distributed storage

Snapshots and images are stored in OSS in the same region. To guarantee service flexibility and resource utilization, when you create a disk from a snapshot or image, the distributed storage does not copy all the data to the new disk at a time. A data block is only copied when it is needed to optimize storage usage. A data block can be less than 10 MiB in size. The next time that you read or write a block, all operations are directly performed on the disk.

Meanwhile, to guarantee the optimal I/O experience, the distributed storage copies the data of a snapshot or image to the disk step by step and block by block via backend duplication when there are less I/O operations.

Therefore, the first time a cloud disk is read or written, its I/O performance may noticeably decrease. However, after the disk has been accessed, its I/O performanc e will return to normal. We recommend that you allow access to the entire disk when you perform a high-load operation, such as reading the disk.

Disk features

A cloud disk has the following features:

- **High data security.**
- **High IOPS for random access and sequential access.**

- Up to 17 disks can be attached to an instance (including system disks and data disks).
- During migration, data is immediately saved before downtime occurs.

2.5.2 Create a disk

This topic describes how to create a new empty disk in the ECS console.

Context

You can create block storage in the console to scale the system storage.

- An ECS instance can have up to 16 data disks (including disks and shared block storage devices).
- A shared block storage device can be attached to a maximum of four ECS instances.
- Each ultra disk (or ultra shared block storage device) and each SSD disk (or SSD shared block storage device) can have a capacity of up to 32 TB.

Note:

- You cannot merge multiple disks in ECS. Each cloud disk that you create is an independent entity. The space on multiple disks cannot be merged through formatting. We recommend that you plan the disk quantity and capacity in advance.
- A snapshot is intended for an independent disk, so data may be different after you perform snapshot rollback under the Logical Volume Management (LVM). We do not recommend that you configure LVM for multiple disks.

Procedure

1. *Log on to the ECS console*.
2. Click the Disks tab. Click Create Disk.
3. On the Create Disk page, configure disk parameters.

   For more information about the parameters, see *Table 2-5: Disk parameters*.

Table 2-5: Disk parameters

Area	Parameter	Description
Region	Region	Required. The region of the disk to be created.

Area	Parameter	Description
	Zone	Required. The zone of the disk to be created.
Basic settings	Name	Required. The name of the disk to be created.
	Department	Required. The department of the disk to be created.
	Project	Required. The project of the disk to be created.
	Storage	Required. Select Disks or Shared Block Storage.
	Type	Required. Select Ultra Disk or SSD Disk and enter the disk size.
	Encrypted	Optional. Indicates whether the new disk is encrypted.
	Use Snapshots	Optional. After selecting `Use Snapshots`, you also need to select the snapshot. **Note:** · After you select Encrypted, this parameter is unavailable. · If the disk size specified is smaller than the selected snapshot size, the actual disk size that is generated will be equal to the snapshot size. Otherwise, the disk size will be equal to the specified value.

4. After you configure the parameters, click OK.

Result

In the disk list, check whether the disk is in the `not attached` state. If yes, the disk is created.

What's next

The procedure varies depending on the operating system of the instance.

- If the Linux operating system is selected for the instance, you must *Attach a disk* and then *Format, partition, and attach data disks in Linux*.

- If the Windows operating system is selected for the instance, you must *Attach a disk* and then *Partition and format data disks in Windows*.

2.5.3 View disks

In the ECS console, you can view disk information.

Procedure

1. *Log on to the ECS console*.
2. Click the Disks tab.
3. Set Department and Region or you can also enter an exact filtering condition. Then click Search.

   Note:
   The filtering condition can be: Disk Name, Disk ID, Disk Status, Disk Usage Type, or Project Name.
4. Click the icon in the Actions column corresponding to the disk and choose View Details from the shortcut menu to go to the Disk Details page.

2.5.4 Roll back a disk

In the ECS console, you can roll back the data on a disk to a previous time.

Prerequisites

The instance of the target disk must be in the `stopped` state.

Notice:
The rollback operation is irreversible. After rollback is finished, the original data cannot be restored. Use caution when calling this operation.

Procedure

1. *Log on to the ECS console*.
2. Click the Snapshots tab. Click the icon in the `Actions` column corresponding to the snapshot and choose Roll Back Disk from the shortcut menu.
3. In the message that appears, click OK.

   Note:

If you select Start Instance Immediately after Rollback, the instance automatically starts after the disk is rolled back.

2.5.5 Modify a disk

In the ECS console, you can modify a disk.

Procedure

1. *Log on to the ECS console*.
2. **On the Disks tab, select the disk to be modified. Click the** icon in the `Actions` **column corresponding to the disk and choose View Details from the shortcut menu to go to the Disk Details page.**

   Note:
   You can also click the disk ID to go to the Disk Details page.
3. **Click Change Properties.**

   You can configure the following parameters:

   - `Disk Name`**: It must be 2 to 128 characters in length and can contain letters, digits, periods (.), underscores (_), and hyphens (-). It must start with an uppercase or lowercase letter.**
   - `Disk Description`**: It must be 2 to 256 characters in length and cannot start with http:// or https://.**
   - `Are you sure you want to configure the disk to delete along with the instance?`: `No` **is selected by default. You can select** `Yes` **if you want to release the disk when the instance is deleted.**
4. **Click OK.**

2.5.6 Attach a disk

2.5.6.1 Overview

After a disk is created, you need to attach the disk. You can only attach independent disks to ECS instances.

Notice:

- **When you attach a disk, check that the ECS instance is in the** `running` **or** `stopped` **state and is not in the** `locked` **state.**

- When you attach a data disk, ensure the disk is in the `not attached` state.
- An ECS instance can have up to 16 data disks (including disks and shared block storage devices).
- An independent cloud disk can be attached only to ECS instances in the same zone, and it is not available for cross-zone attachment.
- An independent disk can be attached to any instance in the same zone and region.

You can attach a disk in the following ways:

- *Attach a disk on the Instance Details page*.
- *Attach a disk on the Disks page*.

2.5.6.2 Attach a disk on the Instance Details page

To attach multiple disks to an instance in the ECS console, you can go to the Instance Details page.

Prerequisites

- Before attaching a disk, you must *Create a disk*.
- When you attach a data disk, ensure the disk is in the `not attached` state.

Context

- You can only attach data disks, but not system disks.
- You do not need to attach data disks that are created at the same time as an instance.
- A disk can only be attached to an instance that is in the same zone and region as the disk.
- An ECS instance can have up to 16 data disks attached. One disk can only be attached to one instance.
- An independent disk can be attached to any instance in the same zone and region .

Procedure

1. *Log on to the ECS console*.
2. On the Instances tab, click the ID of the ECS instance to which the disk is attached.
3. On the Instance Details page, click the Disks tab.

4. Click Attach.
5. In the dialog box, you can configure the following parameters:
   - `Target Disk`: required. Select an existing disk which is in the `not attached` state.
   - `Delete with Instance?`: `No` is selected by default. You can select `Yes` if you want to release the disk when the instance is deleted.
6. Click Submit.

2.5.6.3 Attach a disk on the Disks page

To attach multiple disks to different instances in the ECS console, you can go to the Disks page.

Prerequisites

- Before attaching a disk, you must *Create a disk*.
- When you attach a data disk, ensure the disk is in the `not attached` state.

Context

- You can only attach data disks, but not system disks.
- You do not need to attach data disks that are created at the same time as an instance.
- A disk can only be attached to an instance that is in the same zone and region as the disk.
- An ECS instance can have up to 16 data disks attached. One disk can only be attached to one instance.
- An independent disk can be attached to any instance in the same zone and region .

Procedure

1. *Log on to the ECS console*.
2. Click the Disks tab.
3. Click the icon in the Actions column corresponding to the disk and choose Attach from the shortcut menu.

4. **In the Attach dialog box that appears, you can configure the following parameters:**
   - `Destination Instance`**: the ECS instance to which you want to attach the disk.**
   - `Are you sure you want to configure the disk to delete along with the instance?`**:** `No` **is selected by default. You can select** `Yes` **if you want to release the disk when the instance is deleted.**
5. **Click Submit.**

2.5.7 Partition and format disks

2.5.7.1 Overview

In the ECS console, you can partition and format disks in the Windows or Linux environment.

ECS supports only secondary partitioning of `data disks`**, but not** `system disks` **(either in the Windows or Linux operating system). If you use a third-party tool to perform secondary partitioning of a system disk, you may encounter unknown risks, such as system failure and data loss.**

Note:
Before you attach a data disk, *Create a disk*.

2.5.7.2 Format, partition, and attach data disks in Linux

This topic describes how to partition, format, and attach data disks for Linux instances.

Prerequisites

- *Connect to an instance*.
- *Create a disk* **and** *Attach a disk*.

Procedure

Data disks for the Linux ECS instance are not partitioned or formatted. You can perform the following steps to partition and format data disks:

1. **View the data disks. Before you partition and format the data disks, run the `fdisk -l` command (instead of `df -h`) to view the data disks.**

   The output of the `fdisk -l` command shows information about data disks, such as the information about *`/dev/vdb`*. If *`/dev/vdb`* is not displayed, the ECS instance has no data disks and you do not need to attach data disks.

```
[root@iZ********eZ ~]# fdisk -lDisk /dev/vda: 42.9 GB, 42949672960
 bytes
255 heads, 63 sectors/track, 5221 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk identifier: 0x00078f9c

Device Boot      Start         End      Blocks   Id  System
/dev/vda1   *           1        5222    41940992   83  Linux
Disk /dev/vdb: 21.5 GB, 21474836480 bytes
16 heads, 63 sectors/track, 41610 cylinders
Units = cylinders of 1008 * 512 = 516096 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk identifier: 0x00000000
```

2. **Partition the data disks. Run the `fdisk /dev/vdb` command to partition the data disks. Run the following commands in sequence as prompted:**

   a) `n`: creates a partition.

   b) `p`: creates a primary partition.

   c) `Partition number (1-4)`: the new partition number, an integer that can range from 1 to 4. You can create up to four partitions. In this example, 1 is entered to indicate Partition 1.

   d) `First cylinder`: the start position of the partition. You can select the default value by pressing the Enter key. You can enter a number that can range from 1 to 41610 and press the Enter key. In this example, the default value 1 is used.

   e) `Last cylinder`: the end position of the partition. You can select the default value by pressing the Enter key. You can enter a number that can range from 1 to 11748 and press the Enter key. In this example, the default value is used.

   f) Optional: If you want to create multiple partitions, you can repeat Steps a through e until all four partitions are configured.

   g) Run the `wq` command to start partitioning.

```
[root@iZ********eZ ~]# fdisk /dev/vdb
Device contains neither a valid DOS partition table, nor Sun, SGI or
 OSF disklabel
Building a new DOS disklabel with disk identifier 0x01ac58fe.
Changes will remain in memory only, until you decide to write them.
```

```
After that, of course, the previous content won't be recoverable.

Warning: invalid flag 0x0000 of partition table 4 will be corrected
 by w(rite)

WARNING: DOS-compatible mode is deprecated. It's strongly
 recommended to
         switch off the mode (command 'c') and change display units
 to
         sectors (command 'u').

Command (m for help): n
Command action
   e   extended
   p   primary partition (1-4)
p
Partition number (1-4): 1
First cylinder (1-41610, default 1):
Using default value 1
Last cylinder, +cylinders or +size{K,M,G} (1-41610, default 41610):
Using default value 41610

Command (m for help): wq
The partition table has been altered!
```

3. **View the new partition. Run the `fdisk -l` command to list all the partitions, as shown in *code*. If the command output shows *`/dev/vdb1`*, the partition *`vdb1`* is created.**

```
[root@iZ********eZ ~]# fdisk -lDisk /dev/vda: 42.9 GB, 42949672960
 bytes
255 heads, 63 sectors/track, 5221 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk identifier: 0x00078f9c

Device Boot      Start         End      Blocks   Id  System
/dev/vda1   *           1        5222    41940992   83  Linux

Disk /dev/vdb: 21.5 GB, 21474836480 bytes
16 heads, 63 sectors/track, 41610 cylinders
Units = cylinders of 1008 * 512 = 516096 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk identifier: 0x01ac58fe
Device Boot      Start         End      Blocks   Id  System
/dev/vdb1               1       41610    20971408+  83  Linux
```

4. **Format the new partition. For example, you can run the `mkfs.ext3 /dev/vdb1` command to format the new partition as ext3. The time required for formatting varies depending on the disk size. You can also format the new partition to another file system type. For example, you can run the `mkfs.ext4 /dev/vdb1` command to format it as ext4.**

 **Note:**

Compared with ext2, ext3 only adds the log function. Compared with ext3, ext4 improves some important data structures. ext4 provides better performance and reliability, and more functions.

```
[root@iZ********leZ ~]# mkfs.ext3 /dev/vdb1
mke2fs 1.41.12 (17-May-2010)
Filesystem label=
OS type: Linux
Block size=4096 (log=2)
Fragment size=4096 (log=2)
Stride=0 blocks, Stripe width=0 blocks
1310720 inodes, 5242852 blocks
262142 blocks (5.00%) reserved for the super user
First data block=0
Maximum filesystem blocks=4294967296
160 block groups
32768 blocks per group, 32768 fragments per group
8192 inodes per group
Superblock backups stored on blocks:
    32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632,
 2654208,
    4096000

Writing inode tables: done
Creating journal (32768 blocks): done
Writing superblocks and filesystem accounting information: done

This filesystem will be automatically checked every 25 mounts or
180 days, whichever comes first.  Use tune2fs -c or -i to override.
```

5. **Add partition information. Run the `echo '/dev/vdb1 /mnt ext3 defaults 0 0' >> /etc/fstab` command to add information about the new partition and run the `cat /etc/fstab` command to view partition information.**

   Notice:

   - **This example adds partition information to the ext3 file system. You can add partition information to another type of file systems, such as ext4.**
   - **Ubuntu 12.04 does not support barriers. Therefore, in Ubuntu 12.04, run the `echo '/dev/vdb1 /mnt ext3 barrier=0 0 0' >> /etc/fstab` command to add partition information.**

```
[root@iZ********eZ ~]# echo '/dev/vdb1 /mnt ext3 defaults 0 0' >> /etc
/fstab
[root@iZbp19cdhgdj0aw5r2izleZ ~]# cat /etc/fstab

#
# /etc/fstab
# Created by anaconda on Thu Aug 14 21:16:42 2014
#
# Accessible filesystems, by reference, are maintained under '/dev/
disk'
# See man pages fstab(5), findfs(8), mount(8) and/or blkid(8) for more
 info
```

```
#
UUID=94e4e384-0ace-437f-bc96-057dd64f42ee / ext4 defaults,barrier=0 1
 1
tmpfs                   /dev/shm                tmpfs   defaults
  0 0
devpts                  /dev/pts                devpts  gid=5,mode=620
  0 0
sysfs                   /sys                    sysfs   defaults
  0 0
proc                    /proc                   proc    defaults
  0 0
/dev/vdb1 /mnt ext3 defaults 0 0
```

To attach the data disk to a specific folder (perhaps to store web pages), modify the *`/mnt`* of the preceding command.

6. **Attach the new partition. Run the `mount -a` command to attach all the partitions listed in *`/etc/fstab`*and run the `df -h` command to check the result. If the following information is displayed, the partitions are attached and the new partitions are available for use.**

```
[root@iZ********eZ ~]# mount -a
[root@iZ********eZ ~]# df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/vda1        40G  5.6G   32G  15% /
tmpfs           499M     0  499M   0% /dev/shm
/dev/vdb1        20G  173M   19G   1% /mnt
```

2.5.7.3 Partition and format data disks in Windows

This topic describes how to partition and format the data disks of a Windows instance.

Prerequisites

- *Connect to an instance*.
- *Create a disk* and *Attach a disk*.

Context

The following steps only apply to Windows 2008.

Procedure

Notice:
If your data disks are in the `Offline` state, change them to the `Online` state first before you allocate volume numbers and capacities to the data disks.

1. **In the lower-left corner of the screen, click the Server Manager icon to start Server Manager.**

2. **In the left-side navigation pane of the Server Manager window, choose Storage > Disk Management from the shortcut menu.**
3. **Right-click an empty partition and choose New Simple Volume from the shortcut menu.**
4. **The New Simple Volume Wizard page appears. Click Next.**
5. **Set the size of the simple volume, which is the partition size. The default value is `Maximum disk space`. You can specify the partition size as needed. After you configure the settings, click Next.**
6. **Specify the drive letter, which is listed in the alphabetic order by default. Click Next.**
7. **Format the partition. We recommend that you format the partition using the default settings provided by the wizard. After you configure the settings, click Next to start formatting.**
8. **When the wizard prompts that the partition has been completed, click Finish to close the wizard. The partition is created.**

2.5.8 Resize the system disk

2.5.8.1 Overview

As the business expands, you may need to increase the system disk size. In that case, you can resize the system disk.

When you resize the system disk, note the following risks and limits:

Risks

- **This operation requires you to stop your instance, which means your business is interrupted.**
- **After resizing the system disk, you must redeploy the business runtime environment in the new system disk. This may result in a long business interruption. Use caution when performing this operation.**
- **Custom snapshots are saved after the system disk is resized. However, because the disk ID changes, you can no longer use the custom snapshots of the original system disk to roll back the new system disk. The saved snapshots can still be used to create custom images.**
- **To leave enough snapshot quota for the auto snapshot policy of the new system disk, you can *delete the snapshots that are no longer needed*.**

- **The original system disk is released after you resize the system disk.**

Limits

- **When resizing a system disk, you cannot reduce the disk capacity.**
- **After you resize the system disk, the IP address and MAC address of your instance remain unchanged.**
- **The system disk type cannot be changed.**
- **You cannot resize the system disk in Windows 2003.**

Procedure

If you want to resize a system disk, perform the following steps:

1. *Create a snapshot for a system disk*
2. *Create an image from a snapshot*
3. *Change the system disk*
4. *Set a snapshot policy for a system disk*

2.5.8.2 Create a snapshot for a system disk

If you want to save the data before resizing the system disk, create a snapshot.

Prerequisites

The system disk must be in the `stopped` state.

Context

If you do not want to save the data in the system disk, skip this procedure and proceed to *Change the system disk*. To avoid impact on your business, try to create snapshots during off-peak hours. It requires about 40 minutes to create a 40 GB snapshot for the first time. Therefore, you must leave sufficient time to create a snapshot. When you create a snapshot, check that the system disk has sufficient free space. We recommend that you reserve 1 GB free space. Otherwise, the system may not start properly after the system disk is resized.

Procedure

1. *Log on to the ECS console*.
2. **Click the Instances tab. Set `Department` and `Region`, or select `Instance Name` from the drop-down list and enter a name to filter instances.**

3. Click the instance ID of the system disk, or click the icon in the `Actions` column corresponding to the disk and choose View Details from the shortcut menu to go to the Instance Details page.
4. Click the Disks tab.
5. Locate the system disk and click the icon in the `Actions` column. Choose Create Snapshot from the shortcut menu.
6. In the Create Snapshot dialog box that appears, enter the snapshot name and click OK.

   Note:
   The snapshot name cannot start with auto because auto is reserved as the prefix name of automatic snapshots created by the system.
7. Click the Snapshots tab to view the creation progress and status of the snapshot. When the `Progress` field of a snapshot reaches 100%, it indicates that the snapshot is created.

2.5.8.3 Create an image from a snapshot

To continue using the current operating system and keep its data, you can create an image after the snapshot is created.

Context

- If you do not want to continue using the current operating system or save its data, skip this procedure and proceed to *Change the system disk*.
- If you want to continue using the current system disk, you need to make an image of the current system disk. After the system disk is resized, you can copy all the data to a new environment.
- You can also select an alternative method to create a system disk image. For more information, see *Create a custom image from a snapshot*.

Notice:
When you create an image, check that the system disk has sufficient free space. We recommend that you reserve 1 GB free space. Otherwise, the system may not start properly after the system disk is resized.

Procedure

1. *Log on to the ECS console*.
2. Click the Instances tab. Set `Department` and `Region`, or select `Instance Name` from the drop-down list and enter a name. Click Search to view instance information.
3. Click the icon in the `Actions` column corresponding to the instance and choose View Details from the shortcut menu to go to the Instance Details page.

   Note:
   On the Instances tab, you can also click the instance ID to go to the Instance Details tab.
4. Click the Snapshots tab. Click the icon in the `Actions` column corresponding to the snapshot and choose Create Custom Image from the shortcut menu.
5. In the Create Custom Image dialog box that appears, enter the name and description of the custom image, and then click OK.

   Notice:
   - Remember the image name. You must select the custom image when replacing the system disk.
   - Do not select Add Data Disk Snapshot. When changing the system disk, you cannot select a data disk together.

Result

After the image is created, it is displayed on the Images tab.

2.5.8.4 Change the system disk

Replacing the system disk refers to allocating a new system disk. The system disk ID changes and the original system disk is released.

Prerequisites

- To avoid data loss, back up all the data of the original system disk.
- Ensure that the instance for which you want to change the system disk is in the `stopped` state.

Context

After the system disk is changed,

- Custom snapshots of the original system disk are saved. The automatic snapshot policy becomes invalid and must be reconfigured.
- The system disk ID changes and the original system disk is released.

Procedure

1. *Log on to the ECS console*.
2. Click the Instances tab.
3. Click the icon in the Actions column corresponding to the instance and choose View Details from the shortcut menu.
4. On the Instance Details tab, click Change System Disk.
5. In the Change System Disk dialog box that appears, configure the following parameters:

   Note:
   Before changing the system disk, we recommend that you read the prerequisites and background information.

   - Image Type: If you want to save data in the original system disk, select the custom image created in *Create an image from a snapshot*. Otherwise, select a public image.
   - System Disk: You cannot change the disk type, but can specify a new disk size. The new disk size must be greater than the original disk size. The maximum value is 500 GB.
6. After you configure the parameters, click OK. The system disk is changed.

   Note:
   Go back to the ECS console to view the task status. The process requires about 10 minutes. After the system disk is changed, the instance automatically starts.

2.5.8.5 Set a snapshot policy for a system disk

After you replace a system disk, you must set a snapshot policy for the new system disk if automatic snapshotting is needed.

For more information, see *Configure an automatic snapshot policy*.

2.5.9 Detach a disk

In the ECS console, you can detach a data disk, but not a system disk. Local disks cannot be detached.

Prerequisites

When you detach a data disk, note that:

- **In Windows, you need to log on to the instance and perform Offline operation for the disk through disk management. After the command is executed, you can log on to the console to detach the disk.**

  > **Note:**
  > **To ensure data integrity, we recommend that you suspend read/write operations for all the file systems in this disk. Otherwise, the data that is not read or written may be lost.**

  You can perform Offline operations by using the following steps:

  1. **Right-click Computer and choose Management from the shortcut menu.**
  2. **On the Computer Management page, choose Storage > Disk Management from the shortcut menu.**
  3. **Right-click the disk to be detached and choose Offline from the shortcut menu.**
- **In the Linux system, you need to log on to the instance and run the `umount` command for the disk to be detached. After the command is executed, log on to the ECS console to detach the disk.**

  > **Note:**
  > **If you have enabled automatic attaching to data disk partitions during instance startup, you must delete the attaching information of the data disk partitions from the /etc/fstab file first before detaching the data disk. Otherwise, you cannot connect to the instance after the instance is restarted.**
- **The data disk to be detached is in the `running` state.**

Procedure

1. *Log on to the ECS console*.

2. **Click the Disks tab. Locate the data disk to be detached. Click the icon in the `Actions` column corresponding to the disk and choose Detach from the shortcut menu.**
3. **In the Detach Disk dialog box that appears, click Submit.**

Result

In the disk list, check whether the disk is in the `not attached` state. If yes, the disk has been detached from the instance.

2.6 Images

2.6.1 Overview

An ECS image is a template that contains the software configurations such as the operating system, application server, and application programs of the ECS instance. When creating an instance, you must specify an ECS image. The operating system and software provided by the ECS image are installed in the created instance. You can create a custom image based on a created instance and create more instances based on the custom image.

2.6.2 Select a suitable image

To create an instance, you must select a suitable image.

When you select an image, note that:

- **Select the region and zone.**
- **Select the Linux or Windows operating system.**

  Do not select the Windows operating system when 512 MB memory is selected . Do not select the 32-bit operating system when 4 GB or higher memory is selected.
- **Select the 32-bit or 64-bit operating system.**

When creating an instance, you can select a custom image or public image.

2.6.3 Create a custom image

2.6.3.1 Overview

You can create a custom image, and then use it to create ECS instances or replace the system disk of an ECS instance.

2.6.3.2 Create a custom image from a snapshot

You can create a custom image from a snapshot on the system disk to fully load the operating system and data environment information in the snapshot to the image.

Prerequisites

- **Only system disks, but not data disks, can be used to create custom images.**
- **Ensure the system disk in your instance has available snapshots.**

Procedure

1. *Log on to the ECS console*.
2. **Click the Instances tab. Locate the instance for which you want to create a custom image and click the**  **icon in the** `Actions` **column. Choose Create Custom Image from the shortcut menu.**
3. **In the Create Custom Image dialog box that appears, configure the following parameters:**
   - `Custom Image Name`**: required. It must be 2 to 128 characters in length and can contain letters, digits, underscores (_), periods (.), and hyphens (-). It must start with a letter.**
   - `Custom Image Description`**: optional. It must be 2 to 256 characters in length and cannot start with http:// or https://.**
4. **After you configure the parameters, click OK.**

2.6.3.3 Create a custom image from an instance

By creating a custom image based on an instance, you can fully replicate all disks of the instance. This includes the data on the system disk and data disks.

Prerequisites

To avoid data security risks, delete sensitive data before creating a custom image.

Context

When you create a full image from an instance, each disk of the instance automatically creates a snapshot, and all the snapshots constitute a complete custom image.

Procedure

1. *Log on to the ECS console*.
2. Click the Instances tab. Locate the instance for which you want to create a custom image and click the icon in the `Actions` column. Choose Create Custom Image from the shortcut menu.
3. In the Create Custom Image dialog box that appears, configure the following parameters:
   - `Custom Image Name`: required. It must be 2 to 128 characters in length and can contain letters, digits, underscores (_), periods (.), and hyphens (-). It must start with a letter.
   - `Custom Image Description`: optional. It must be 2 to 256 characters in length and cannot start with http:// or https://.
4. After you configure the parameters, click OK.

2.6.4 View images

In the ECS console, you can view image information.

Procedure

1. *Log on to the ECS console*.
2. Click the Images tab. Set `Department` and `Region`, or select `Image Name` from the drop-down list and enter a name. Then, click Search to view image information.

> **Note:**
> You can click Image Name to select other filtering conditions from the drop-down list: `Image ID` and `Image Type`.

2.6.5 Copy a custom image

You can copy a custom image from one region to another. This operation is applicable to scenarios where applications are deployed across regions and where ECS instances in different regions run the same image environment.

Context

The time required to copy an image depends on the network status and concurrent tasks in the queue.

Procedure

1. *Log on to the ECS console*.
2. Click the Images tab. Click the icon in the `Actions` column corresponding to the snapshot and choose Copy Image from the shortcut menu.
3. In the Copy Image dialog box that appears, you can find the ID of the selected image. You can configure the following parameters:
   - Name: required. The name of the custom image shown in the target region.
   - Description: optional. The description of the custom image shown in the target region.

     Note:
     It must be 2 to 256 characters in length and cannot start with http:// or https://.
4. Click OK.
5. Switch to the target region and you can see that the custom image is in the `creating` state. When the status becomes `Available`, the copy process is completed.

2.6.6 Share custom images

You can share your custom images with other departments.

Prerequisites

Only custom images can be shared.

Procedure

1. *Log on to the ECS console*.
2. Click the Images tab. Click the icon in the `Actions` column corresponding to the snapshot and choose Share Image from the shortcut menu.
3. In the dialog box that appears, select `Department` and click OK.

2.6.7 Import a custom image

2.6.7.1 Overview

In the ECS console, you can import a custom image. You can create instances by using a custom image, which may be created from your local server, virtual machines, or a cloud server of other cloud platforms.

2.6.7.2 Limits on importing custom images

This topic describes limits on importing images to ensure Image availability and improve import efficiency.

Limits on importing custom images:

- *Limits on importing custom images in Linux*
- *Limits on importing custom images in Windows*

Limits on importing custom images in Linux

When you import custom images in Linux, note the following limits:

- **Multiple network interfaces are not supported.**
- **IPv6 addresses are not supported.**
- **The password must be 8 to 30 characters in length. It must contain uppercase and lowercase letters, digits, and special characters.**
- **The firewall is disabled, and port 22 is enabled by default.**
- **The Linux system disk size is between 40 GiB and 500 GiB.**
- **DHCP must be enabled in the image.**
- **SELinux is disabled.**
- **The Kernel-based Virtual Machine (KVM) driver must be installed.**
- **We recommend that you install cloud-init to configure the hostname and NTP and yum sources.**
- **The imported Red Hat Enterprise Linux (RHEL) image must have a BYOL license.**

Table 2-6: Limits

Item	Standard operating system image	Non-standard operating system image
Definition	Supported standard operating systems (both 32-bit and 64-bit) include: · CentOS · Ubuntu · SUSE · OpenSUSE · Red Hat · Debian · CoreOS · Aliyun Linux	Non-standard operating systems include: · Operating systems that are not supported by Alibaba Cloud. · Standard operating systems that do not meet the requirements of critical system configuration files, basic system environments, and applications. If you want to use non-standard operating system images, you must select Others Linux when importing images. If you import non-standard operating system images, Alibaba Cloud does not pre-configure the created instances. After you create an instance, you must connect to the instance by clicking Connect to Management Terminal in the ECS console. You can then configure the IP address, route, and password.

Item	Standard operating system image	Non-standard operating system image
Critical system configuration files	· Do not modify `/etc/issue*`. Otherwise, the system release cannot be identified, leading to system creation failure. · Do not modify `/boot/grub/menu.lst`. Otherwise, the system may fail to start. · Do not modify `/etc/fstab`. Otherwise, partitions cannot be loaded, leading to system startup failure. · Do not modify `/etc/shadow` to read-only. Otherwise, the password file cannot be modified, leading to system creation failure. · Do not modify `/etc/selinux/config` to enable SELinux. Otherwise, the system may fail to start.	Requirements for standard operating system images are not met.
Basic system environments	· Do not adjust the system drive partition. Only a single root partition is supported. · Ensure that the system disk has a sufficient free space. · Do not modify critical system files, such as `/sbin`, `/bin`, and `/lib*`. · Before importing an image, confirm the integrity of the file system. · Only file systems ext3 and ext4 are supported.	

Item	Standard operating system image	Non-standard operating system image
Applications	Do not install `qemu-ga` in a user-defined image. Otherwise, some of the services that Alibaba Cloud requires may be unavailable.	
Image file formats	Only RAW and VHD images can be imported. If you want to import images in other formats, use a tool to convert the format before importing the image. We recommend that you import images in VHD format which has a smaller transmission capacity.	
Image file size	We recommend that you configure the disk size for importing images based on the virtual disk size (not the image file size). The disk size for importing images must be greater than or equal to 40 GiB.	

Limits on importing custom images in Windows

When you import custom images in Windows, note the following limits:

- **The password must be 8 to 30 characters in length. It must contain uppercase and lowercase letters, digits, and special characters.**
- **Imported Windows images do not provide the Windows activation service.**
- **The firewall must be disabled. Otherwise, remote logon is unavailable. Port 3389 must be enabled.**
- **The Windows system disk size is between 40 GiB and 500 GiB.**

Table 2-7: Limits

Item	Description
Operating system versions	**Alibaba Cloud supports importing the following versions of operating system images (32-bit and 64-bit):** **· Microsoft Windows Server 2016** **· Microsoft Windows Server 2012, including:** **- Microsoft Windows Server 2012 R2 (Standard Edition)** **- Microsoft Windows Server 2012 (Standard Edition and Datacenter Edition)** **· Microsoft Windows Server 2008, including:** **- Microsoft Windows Server 2008 R2 (Standard Edition, Datacenter Edition, and Enterprise Edition)** **- Microsoft Windows Server 2008 (Standard Edition, Datacenter Edition, and Enterprise Edition)** **· Microsoft Windows Server 2003, including:** **- Microsoft Windows Server 2003 R2 (Standard Edition, Datacenter Edition, and Enterprise Edition)** **- Microsoft Windows Server 2003 (Standard Edition, Datacenter Edition, and Enterprise Edition) or later versions, including Service Pack 1 (SP1)** **· Microsoft Windows 7, including:** **- Microsoft Windows 7 (Professional Edition)** **- Microsoft Windows 7 (Enterprise Edition)**
Requirements for the basic system environment	**· Multi-partition system disks are supported.** **· Ensure that the system disk has a sufficient free space.** **· Do not modify critical system files.** **· Before importing an image, confirm the integrity of the file system.** **· The NTFS file system with the MBR partition type is supported.**

Item	Description
Applications	Do not install qemu-ga in an imported image. If it is installed, some of the services that Alibaba Cloud needs may be unavailable.
Image file formats	· RAW · VHD We recommend that you configure the system disk size for importing images based on the virtual disk size (not the image file size). The system disk size for importing images must be between 40 GiB and 500 GiB. Note: We recommend that you import images in VHD format which has a smaller transmission capacity.

2.6.7.3 Convert the image file format

Only image files in RAW or VHD format can be imported. If you want to import images in other formats, convert the format before importing the image.

Prerequisites

You can use the qemu-img tool to convert image files into VHD or RAW from other formats, such as RAW, Qcow2, VMDK, VDI, VHD (vpc), VHDX, qcow1, or QED. You can also use qemu-img to convert image files between RAW and VHD formats.

Install qemu-img and convert image file format

You can use different methods to install qemu-img and convert the image file format based on the operating system of your computer:

- *Windows*
- *Linux*

Windows

To install qemu-img and convert the image file format, perform the following steps:

1. **Download and** *install qemu* **to** `C:\Program Files\qemu`**.**

2. **Configure environment variables as follows:**
   a. **Choose Start > Computer and right-click Properties.**
   b. **In the left-side navigation pane, click Advanced System Settings.**
   c. **In the System Properties dialog box that appears, click the Advanced Tab and click Environment Variables.**
   d. **In the Environment Variables dialog box that appears, find in the System variables list the `Path` variable, and then click Edit. If the `Path` variable does not exist, click New.**
   e. **Add a variable value:**
      - **In the `Variable value` field of the Edit System Variable dialog box that appears, enter *`C:\Program Files\qemu`*. Separate multiple values with semicolons (;).**
      - **In the `Variable name` field of the New System Variable dialog box that appears, enter *`Path`*. In the `Variable value` field, enter *`C:\Program Files\qemu`*. Click OK.**
3. **Open Command Prompt in Windows and run the `qemu-img --help` command. If a successful prompt is displayed, the installation is completed.**
4. **In Command Prompt, run the `cd [Directory of the source image file]` command to change the file directory, such as `cd D:\ConvertImage`.**
5. **Run the following command in Command Prompt to convert the image file format: `qemu-img convert -f raw -O qcow2 centos.raw centos.qcow2`.**

   The command parameters are described as follows:

   - **`-f` is followed by the source image format.**
   - **`-O` (uppercase is required) is followed by the converted image format, the source file name, and the target file name.**

   After the file format is converted, the target file appears in the directory of the source image file.

Linux

To install qemu-img and convert the image file format, perform the following steps:

1. Install qemu-img:
   - For Ubuntu, run the `apt install qemu-img` command.
   - For CentOS, run the `yum install qemu-img` command.
2. Run the following command to convert the image file:

   ```
   qemu-img convert -f raw -O qcow2 centos.raw centos.qcow2
   ```

   The command parameters are described as follows:

   - `-f` is followed by the source image format.
   - `-O` (uppercase is required) is followed by the converted image format, the source file name, and the target file name.

2.6.8 Export a custom image

In the ECS console, you can export an existing custom image.

Prerequisites

Ensure that you are authorized to export images. For more information about the authorization process, see RAM management in *Apsara Stack Console User Guide* .

Procedure

1. *Log on to the ECS console*.
2. Click the Images tab. Click the icon in the `Actions` column corresponding to the snapshot and choose Export Image from the shortcut menu.
3. In the dialog box that appears, select `OssBucket` and enter `OSS Prefix`. Click OK.

   Note:
   `OSS Prefix` is an optional parameter. It must be 1 to 30 characters in length and can contain digits and letters.

2.6.9 Delete an image

In the ECS console, you can delete an image that is no longer needed.

Prerequisites

You cannot delete public images.

Procedure

1. *Log on to the ECS console*.

2. Click the Images tab. Set `Department` and `Region`, or select `Image Name` from the drop-down list and enter a name. Click Search to find the image to be deleted.
3. Click the icon in the `Actions` column corresponding to the snapshot and choose Delete from the shortcut menu.

   Note:
   Select multiple images and then click Delete in the upper-right corner to delete multiple images at a time.
4. In the message that appears, click OK.

2.7 Snapshots

2.7.1 Overview

A snapshot saves the state of disk data at a certain point in time for later data backup or custom image creation.

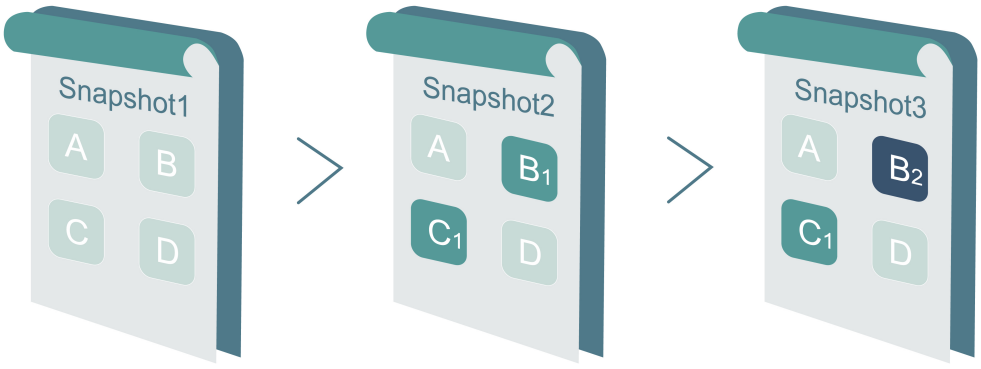
When you use disks, note that:

- When writing or saving data to a disk, you can use the data on another disk as the basic data of this disk.
- Although the disk provides a secure storage mode, make sure that complete data is stored. However, if the data stored on the disk is incorrect (for example, due to an application error, or a vulnerability in the application was exploited), a mechanism is required. This mechanism is to ensure that your data can be stored to the desired state when a problem occurs.

Alibaba Cloud allows you to create a snapshot to save a copy of the data on a disk at a specific point in time. You can create disk snapshots on a scheduled basis to guarantee continuous operations of your business.

Snapshots use an incremental method. Two snapshots are compared and only the changed data is copied, as shown in *Figure 2-7: Snapshot operation guide*.

Figure 2-7: Snapshot operation guide



Snapshot 1, Snapshot 2, and Snapshot 3 are the first, second, and third disk snapshots. The file system checks the disk data block by block. When a snapshot is created, only the blocks with changed data will be copied to the snapshot.

Because Snapshot 1 is the first disk snapshot, it copies all of the data on the disk. Snapshot 2 only copies the changed data blocks B1 and C1 and references blocks A and D in Snapshot 1 as its data blocks A and D. Snapshot 3 copies the changed data block B2, and references blocks A and D in Snapshot 1 as its data blocks A and D. Then, Snapshot 3 references data block C1 in Snapshot 2 as its data block C1.

If you want to restore the disk to the status at the time of Snapshot 3, you can perform snapshot rollback. Rollback is used to copy data blocks A, B2, C1, and D in Snapshot 3 to the disk.

If Snapshot 2 is deleted, data block B1 in the snapshot is deleted but data block C1 is not. In this way, when you restore the disk to the status at the time of Snapshot 3, data block C1 can also be restored.

Snapshots are stored in your Object Storage Service (OSS) buckets, but the OSS console does not allow you to query, manage, or calculate bucket space used by snapshot files. You can only operate snapshots by using the ECS console or APIs.

2.7.2 Create a snapshot

In the ECS console, you can create a snapshot of a disk as needed.

Context

You can create a limited number of snapshots of a disk. You can create up to 64 snapshots of each disk.

Procedure

1. *Log on to the ECS console*.
2. Click the Disks tab. Locate the snapshot to be created and click the icon in the `Actions` column. Choose Create Snapshot from the shortcut menu.
3. Enter the snapshot name and description and then click OK.
4. Click the Snapshots tab to view the creation progress and status of the snapshot. When the `Progress` field of a snapshot reaches 100%, it indicates that the snapshot is created.

> **Note:**
>
> - The first time you create a full snapshot of a disk requires a long time.
> - Otherwise, you can create another snapshot of a disk (with existing snapshots) and it only requires a short time. The duration depends on the volume of data that is changed between the most latest and previous snapshots. The more data that is changed, the longer the duration.
> - Avoid creating snapshots during peak business hours.

2.7.3 View snapshots

In the ECS console, you can view snapshot information.

Procedure

1. *Log on to the ECS console*.
2. Click the Snapshots tab. Set `Department` and `Region`, or select `Snapshot Name` from the drop-down list and enter a name. Click Search to view snapshot information.

> **Note:**
>
> You can click Snapshot Name to select other filtering conditions from the drop-down list: `Snapshot ID`, `Disk Name`, `Project Name` and `Created At`.

2.7.4 Delete a snapshot

In the ECS console, you can delete a snapshot that is no longer needed.

Prerequisites

- **Deleted snapshots cannot be restored. Use caution when performing this operation.**
- **If a system disk snapshot has been used to create a custom image, the snapshot cannot be deleted.**

Procedure

1. *Log on to the ECS console*.
2. **Click the Snapshots tab. Set Department and Region or you can also enter an exact filtering condition. Then click Search.**

   Note:
   The filtering condition can be: Snapshot Name, Snapshot ID, Disk Name, Project Name, or Created At.
3. **Click the icon in the Actions column corresponding to the snapshot and choose Delete from the shortcut menu.**

   Note:
   Select multiple snapshots and then click Delete in the upper-right corner to delete multiple snapshots at a time.
4. **In the message that appears, click OK.**

2.7.5 Scenarios

This topic describes how to use snapshots to create images and data disks.

In addition to rolling back the source disks, you can also use snapshots to:

- **Create custom images.**
- **Create data disks when you create an instance.**

Create a custom image

You can create a custom image to quickly replicate an existing instance. For more information about the procedure, see *Create a custom image from a snapshot*.

Notice:
You cannot use a data disk snapshot to create custom images.

Create data disks for an instance

You can use a snapshot to create a data disk. The new data disk will contain the same data as the data disk of another instance.

Procedure

When you *Create a disk*, **select** `Use Snapshots` **to create a data disk by using the snapshot of another data disk in the same region. The capacity of the new data disk is determined by the snapshot capacity and cannot be modified.**

Notice:
If you reset a data disk that was created from a snapshot, the data in the snapshot is restored to the data disk.

2.8 Automatic snapshot policies

2.8.1 Overview

You can specify a time to create a daily snapshot in a day, on which days in a week to create snapshots, and how long snapshots are saved. You can apply an automatic snapshot policy to disks based on your business needs.

2.8.2 Create an automatic snapshot policy

In the ECS console, you can create an automatic snapshot policy as needed.

Procedure

1. *Log on to the ECS console*.
2. **Click the Snapshot Policies tab. Click Create.**

3. **Configure automatic snapshot policy parameters.**

   For more information about the parameters, see *Table 2-8: Automatic snapshot policy parameters*.

   Table 2-8: Automatic snapshot policy parameters

Parameter	Description
Name	The name of the automatic snapshot policy. It must be 2 to 128 characters in length. It can contain letters, digits, underscores (_), and hyphens (-). It must start with an uppercase or lowercase letter.
Region	The region of the automatic snapshot policy.
Department	The department of the automatic snapshot policy.
Time	The time when the snapshot is created. You can select multiple values. The time ranges from 00:00 to 23:00.
Recurrence	The day when the snapshot is created. You can select multiple values. The day ranges from Monday to Sunday.
Snapshot Retention Period (Days)	Valid values: 1 to 65,535. The snapshot is saved forever by default.

4. **After you configure the parameters, click OK.**

What's next

After the automatic snapshot policy is created, you must *Configure an automatic snapshot policy*.

2.8.3 View automatic snapshot policies

In the ECS console, you can view automatic snapshot policy information.

Procedure

1. *Log on to the ECS console*.

2. Click the Snapshot Policies tab. Set `Region`, or select `Policy ID` from the drop-down list and enter an ID. Click Search to view automatic snapshot policy information.

> **Note:**
> You can click Policy ID to select other filtering conditions from the drop-down list: `Automatic Snapshot Policy Name`.

2.8.4 Modify an automatic snapshot policy

In the ECS console, you can modify configurations for automatic snapshot policies.

Procedure

1. *Log on to the ECS console*.
2. Click the Snapshot Policies tab. Click the icon in the `Actions` column corresponding to an automatic snapshot policy and choose Change from the shortcut menu.
3. In the dialog box that appears, set the `Policy Name`, `Time`, `Recurrence`, and `Snapshot Retention Period` parameters as needed.
4. Click OK.

2.8.5 Configure an automatic snapshot policy

In the ECS console, you can configure an automatic snapshot policy for a disk.

Prerequisites

- The disk must be in the `running` state.
- The automatic snapshot is named in the following format: auto_yyyyMMdd_1, such as auto_20140418_1.

> **Note:**
> - We recommend that the automatic snapshot task is performed during off-peak hours.
> - Manually created snapshots do not conflict with automatic snapshots. You can only create a manual snapshot after an automatic snapshot is complete.

Procedure

1. *Log on to the ECS console*.

2. **Click the Disks tab. Locate the disk where you want to configure the automatic snapshot policy and click the icon in the `Actions` column. Choose Apply Snapshot Policy from the shortcut menu.**
3. **In the Apply Snapshot Policy dialog box, select `Automatic Snapshot Policy` and then click OK.**

2.8.6 Configure an automatic snapshot policy for multiple disks

In the ECS console, you can configure automatic snapshot policies for multiple disks at a time.

Procedure

1. *Log on to the ECS console*.
2. **Click the Snapshot Policies tab. Click the icon in the Actions column corresponding to the automatic snapshot policy and choose Configure from the shortcut menu.**
3. **Select one or many disks to apply the automatic snapshot policy.**
   - **Select multiple disks in the Available Disks list and then click → to select multiple disks.**
   - **Select multiple disks in the Selected Disks list and then click ← to clear multiple disks.**
   - **Click Select All in the upper-right corner of the Available Disks list and then click → to select all disks.**
   - **Click Select All in the upper-right corner of the Selected Disks list and then click ← to clear all disks.**
4. **Click OK.**

2.8.7 Delete an automatic snapshot policy

In the ECS console, you can delete an automatic snapshot policy that is no longer needed.

Procedure

1. *Log on to the ECS console*.
2. **Click the Snapshot Policies tab. Set `Region`, or select `Policy ID` or `Automatic Snapshot Policy Name` from the drop-down list and enter an ID or a name. Click Search to find the specified snapshot policy.**

3. Click the icon in the `Actions` column corresponding to the automatic snapshot policy and choose Delete from the shortcut menu.
4. Click OK.

2.9 Security groups

2.9.1 Overview

A security group is a virtual firewall that provides Stateful Packet Inspection (SPI). A security group is used for network access control for one or many cloud servers. As an important method of network security isolation, it is used to divide security domains on the cloud.

Security group limits

- A single security group cannot contain more than 1,000 instances. If more than 1,000 instances need access to each other, you can allocate them to different security groups and permit mutual access through mutual authorization.
- Each instance can join a maximum of five security groups.
- Each user can have a maximum of 100 security groups.
- Each security group can have a maximum of 100 security group rules.
- Modifying security groups will not affect your services.
- Security groups are stateful. If outbound packets are allowed over a connection, inbound packets over this connection are also allowed.

Security group rules

Security group rules can allow or deny inbound or outbound traffic for ECS instances associated with the security groups.

You can authorize or cancel security group rules at any time. Changes in security group rules are automatically applied to ECS instances associated with the security group.

Security group rules must be precise. If you associate an instance with multiple security groups, hundreds of rules may apply to the instance, which may cause the instance to become inaccessible.

2.9.2 View security groups

In the ECS console, you can view security group information.

Procedure

1. *Log on to the ECS console*.
2. **Click the Security Groups tab. Set Department and Region or you can enter an exact filtering condition. Then click Search.**

   Note:
   The filtering condition can be: Security Group Name, Security Group ID, VPC ID, or Project Name.
3. **Click the icon in the Actions column corresponding to the security group and choose View Details from the shortcut menu to view more details of instances and rules in the security group.**

   Note:
   You can click the security group ID to view more details.

2.9.3 Add security group rules

You need to add appropriate rules to a security group as needed.

Procedure

1. *Log on to the ECS console*.
2. **Click the Security Groups tab. Set** `Department` **and** `Region`**, or select** `Security Group Name` **from the drop-down list and enter a name. Then click Search to find the specified security group.**
3. **Click the security group ID. The ECS Instances tab is displayed by default.**
4. **Click the Rules tab. Click Create Security Group Rule.**

5. **In the dialog box that appears, configure the parameters of the security group and click OK.**

   The security group parameters are as follows:

   - `Authorization Policy`**: Select** `Allow` **or** `Block`**.**

     Block policies discard the data packet without returning a response. If two security groups have the same rules but different authorization policies, `Block` **policies are used while** `Allow` **policies are ignored.**

   - `Rule Direction`**:**
     - **Outbound: Your ECS instances access other ECS instances in the internal network or resources on the public network.**
     - **Inbound: Other ECS instances in the internal network or resources in the public network access your ECS instances.**
   - `Protocol Type` **and** `Port Range`**: The port range is based on the protocol type.** *Table 2-9: Parameter description* **lists the relationship between protocol types and port ranges.**

     Table 2-9: Parameter description

Protocol type	Port range	Scenario
All	-1/-1 is displayed, indicating all ports. You cannot set a port range for this protocol type.	It is used in all trusted scenarios.
TCP	The port range can be customized. Valid values: 1 to 65535. It is in the format of Start Port/End Port. For a single port, you need to set the port range in the standard format, such as 80/80 to represent port 80.	It can be used to allow or block one or several successive ports.
UDP		
ICMP	-1/-1 is displayed, indicating all ports. You cannot set a port range for this protocol type.	You can run the `ping` command to check network connection status between instances.

Protocol type	Port range	Scenario
GRE	-1/-1 is displayed, indicating all ports. You cannot set a port range for this protocol type.	It is used for VPN.

- `Priority`: the priority of the rule. The default value is 1, which indicates the highest priority.
- `Authorization Type` and `Authorized IP address`: The authorized IP address is based on the authorization type. *Table 2-10: Authorization parameter description* lists the relationship between authorization types and authorized IP addresses.

  Table 2-10: Authorization parameter description

Authorization Type	Authorized IP Address
IP Address Range Access	Enter an IP address or CIDR block, in the format of 12.1.1.1 or 13.1.1.1/25. Only IPv4 addresses are supported. 0.0.0.0/0 indicates whether to allow or deny all IP addresses. Use caution when setting 0.0.0.0/0.
Security Group Access	Security Group Access is only applicable to internal network access. IP Address Range Access must be selected for public network access.

- `Description`: the description of the rule for later management.

2.9.4 Remove an instance from a security group

In the ECS console, you can remove an instance from a security group.

Procedure

1. *Log on to the ECS console*.
2. Click the Security Groups tab. Set `Department` and `Region`, or select `Security Group Name` from the drop-down list and enter a name. Then click Search to find the specified security group.
3. Click the security group ID to go to the ECS Instances page.
4. Click the icon in the Actions column corresponding to the security group and choose Remove from Security Group from the shortcut menu. In the message that appears, click OK.

2.9.5 Delete a security group

In the ECS console, you can delete a security group that is no longer needed.

Prerequisites

All instances must be removed from the security group. Otherwise, the security group cannot be deleted.

Procedure

1. *Log on to the ECS console*.
2. **Click the Security Groups tab. Set** `Department` **and** `Region`**, or select** `Security Group Name` **from the drop-down list and enter a name. Then click Search to find the specified security group.**
3. **Click the**  **icon in the** `Actions` **column corresponding to the security group and choose Delete from the shortcut menu.**

    **Note:**
   On the Security Groups tab, you can select multiple security groups and then click Delete. In the dialog box that appears, select `Yes` **and then click OK.**
4. **In the message that appears, click OK.**

2.10 ENIs

2.10.1 Overview

Elastic network interfaces (ENIs) are divided into primary ENIs and secondary ENIs. The primary ENI is created by default when an instance is created in a VPC. The lifecycle of the primary ENI is the same as that of the instance and you cannot detach the primary ENI from the instance. A secondary ENI is created separately. You can attach it to or detach it from an instance. This topic describes secondary ENIs.

2.10.2 Create an ENI

In the ECS console, you can create an ENI.

Procedure

1. *Log on to the ECS console*.
2. **Click the Elastic Network Interfaces (ENIs) tab. Click Create ENI.**

3. **In the Create ENI dialog box, configure the parameters of the NIC and click OK.**

   For more information about the parameters, see *Table 2-11: ENI parameters*.

   Table 2-11: ENI parameters

Area	Parameter	Description
Region	Region	Required. The region of the ENI to be created.
	Zone	Required. The physical zones in the same region with separate power supplies and networks. Zones in the same region can access each other, but faults in one zone are limited to the zone. We recommend that you create instances in different zones to improve application availability.
Description	Department	Required. The department of the ENI to be created.
	Project	Required. The project of the ENI to be created.
	VPC	Required. When you attach an ENI to an instance, they must be in the same VPC. **Note:** After an ENI is created, you cannot change its VPC.
	Security Groups	Required. The security group of the current VPC.
	ENI Name	Optional. The name of the ENI.
	IP Address	Optional. The primary IPv4 address of the ENI. The IPv4 address must be available in the CIDR block of the specified VSwitch. If you do not specify one, a private IP address is automatically assigned to your ENI after it is created.
	Description	Description: The description of the ENI for later management.

2.10.3 View ENIs

In the ECS console, you can view ENI information.

Procedure

1. *Log on to the ECS console*.
2. **Click the Elastic Network Interfaces (ENIs) tab. Set** `Department` **and** `Region`**, or select** `NIC ID` **from the drop-down list and enter an ID. Then click Search to find the specified ENI and view its information.**

> **Note:**
> **You can click NIC ID to select other filtering conditions from the drop-down list:** `NIC name`, `Project Name`, `VPC ID`, `VSwitch ID`, **and** `Instance ID`.

2.10.4 Modify an ENI

In the ECS console, you can modify ENI information.

Procedure

1. *Log on to the ECS console*.
2. **Click the Elastic Network Interfaces (ENIs) tab. Locate the secondary ENI to be modified and click the** icon **in the** `Actions` **column. Choose Change from the shortcut menu.**
3. **In the dialog box that appears, modify the** `NIC Name`,`Security Groups`, **or** `Description` **of the ENI and click OK.**

2.10.5 Attach an ENI to an instance

After creating an ENI, you can attach it to an instance.

Prerequisites

Note the following limits when you attach an ENI to an instance:

- **You can only attach a secondary ENI to an instance.**
- **You have created an ENI as described in** *Create an ENI*. **The ENI must be in the** `available` **state.**
- **The instance must be in the running or stopped state. For more information about how to start and stop an instance, see** *Stop, restart, or start an instance*.
- **The ENI must be in the same VPC as the instance.**

- The VSwitches of the ENI and the instance can be different, but they must be in the same zone.
- An ENI can only be attached to one ECS instance at a time. However, an ECS instance can have multiple ENIs. For more information about the maximum number of ENIs that can be attached to each instance type, see Instance types in *ECS Product Introduction* .

Procedure

1. *Log on to the ECS console*.
2. Click the Elastic Network Interfaces (ENIs) tab. Locate the secondary ENI to be attached and click the  icon in the `Actions` column. Choose Attach to ECS Instance from the shortcut menu.
3. In the dialog box that appears, select a `Destination Instance` and click OK.

2.10.6 Detach an ENI from an instance

In the ECS console, you can detach ENIs from instances. You can only detach secondary ENIs, but not primary ENIs.

Prerequisites

- The secondary ENI must be in the Bound state.
- The instance must be in the Running or Stopped state. For more information about how to start and stop an instance, see *Stop, restart, or start an instance*.

Procedure

1. *Log on to the ECS console*.
2. Click the Elastic Network Interfaces (ENIs) tab. Find the secondary ENI to be detached and click the  icon in the `Actions` column. Choose Detach from ECS Instance from the shortcut menu.
3. In the message that appears, click OK.

2.10.7 Delete an ENI

In the ECS console, you can delete an ENI that is no longer needed.

Context

You can only delete ENIs one by one.

Prerequisites

The ENI must be in the `available` state.

Procedure

1. *Log on to the ECS console*.
2. **Click the Elastic Network Interfaces (ENIs) tab. Set `Department` and `Region`, or select `NIC ID` from the drop-down list and enter an ID. Then click Search to find the ENI to be deleted.**
3. **Click the icon in the `Actions` column corresponding to the ENI and choose Delete from the shortcut menu.**
4. **In the message that appears, click OK.**

2.11 Deployment sets

2.11.1 Overview

A deployment set allows you to view the physical topology of hosts, racks, and switches. You can also select a deployment policy that best suits the reliability and performance requirements of your business.

2.11.2 Create a deployment set

In the ECS console, you can create a deployment set as needed.

Procedure

1. *Log on to the ECS console*.
2. **Click the Deployment Sets tab. Click Create Deployment Set.**
3. **In the Create Deployment Set dialog box, configure the parameters of the deployment set.**

   For more information about the parameters, see *Table 2-12: Deployment set parameters*.

   Table 2-12: Deployment set parameters

Area	Parameter	Description
Region	Region	Required. The region of the deployment set to be created.

Area	Parameter	Description
	Zone	Required. The physical zones in the same region with separate power supplies and networks. Zones in the same region can access each other, but faults in one zone are limited to the zone. We recommend that you create instances in different zones to improve application availability.
Description	Department	Required. The department of the deployment set to be created.
	Project	Required. The project of the deployment set to be created.
	Deployment Domain	Required. You can select Default or Switch.
	Deployment Granularity	Required. You can select Host Machine, Rack , or Switch.
	Deployment Policy	Required. You can select Loose Dispersion or Strict Dispersion.
	Deployment Set Name	Optional. The name of the deployment set.
	Description	Optional. The description of the deployment set.

4. After you configure the parameters, click OK.

Result

After the deployment set is created, you can view its parameters, which are consistent with those you configured.

2.11.3 View a deployment set

In the ECS console, you can view deployment set information.

Procedure

1. *Log on to the ECS console*.
2. Click the Deployment Sets tab. Set `Department` and `Region`, or select `Deployment Set Name` from the drop-down list and enter a name. Then click Search. The

search results show a list of deployment sets based on your search criteria, and you can view more details.

> **Note:**
> You can click Deployment Set Name to select other filtering conditions from the drop-down list: `Deployment Set ID` and `Project Name`.

2.11.4 Modify a deployment set

In the ECS console, you can modify deployment set information.

Procedure

1. *Log on to the ECS console*.
2. Click the Deployment Sets tab. Click the icon in the `Actions` column corresponding to a configuration set and choose Change from the shortcut menu.
3. In the dialog box that appears, modify the `Name` or `Description` of the deployment set and click OK.

2.11.5 Delete a deployment set

In the ECS console, you can delete a deployment set that is no longer needed.

Prerequisites

ECS instances must have been completely removed from the deployment set.

Procedure

1. *Log on to the ECS console*.
2. Click the Deployment Sets tab. Click the icon in the `Actions` column corresponding to a configuration set and choose Delete from the shortcut menu.
3. In the message that appears, click OK.

2.12 Install FTP software

2.12.1 Overview

File Transfer Protocol (FTP) transfers files between a client and a server by establishing two TCP connections. One is the command link for transferring commands between a client and a server. The other is the data link used to upload

or download data. Before uploading files to an instance, you must build an FTP site for the instance.

2.12.2 Install VSFTP in CentOS

This topic describes how to install VSFTP in CentOS before uploading files.

Procedure

1. **Install VSFTP. Run the `yum install vsftpd -y` command to install VSFTP.**
2. **Add an FTP account and directory.**
   a. **Check the *nologin* directory, which is */usr/sbin/nologin* or */sbin/nologin*.**
   b. **Create an FTP account. Run the */alidata/www/wwwroot* command to specify the directory of user pwftp, or run the `useradd -d /alidata/www/wwwroot -s /sbin/nologin pwftp` command to specify an FTP account and its directory.**
   c. **Run the following command to change the account password:**

      ```
      passwd pwftp
      ```

   d. **Run the following command to modify the permissions on the specified directory:**

      ```
      chown -R pwftp.pwftp /alidata/www/wwwroot
      ```

3. **Configure VSFTP.**
   a. **Run the `vi /etc/vsftpd/vsftpd.conf` command to modify the VSFTP configuration file.**
   b. **Modify `anonymous_enable=YES` to `anonymous_enable=NO`.**
   c. **Uncomment the following configuration lines: `#`:**

      ```
      local_enable=YES write_enable=YES chroot_local_user=YES
      ```

   d. **To save the modifications and exit, press the ESC key and enter `wq`.**
4. **Modify shell configurations. Use the `vi` editor to modify */etc/shells*. If the file does not contain */usr/sbin/nologin* or */sbin/nologin* (depending on the current system configurations), add it to the file.**
5. **Start VSFTP and perform a logon test.**
   a. **Run the `service vsftpd start` command to start VSFTP.**
   b. **Use the pwftp account to perform a FTP logon test. The directory is */alidata/www/wwwroot*.**

2.12.3 Install VSFTP in Ubuntu or Debian

This topic describes how to install VSFTP in Ubuntu or Debian.

Procedure

1. **Run the `apt-get install vsftpd -y` command to install VSFTP.**
2. **Add an FTP account and directory.**
   a. **Check the nologin directory, which is *`/usr/sbin/nologin`* or *`/sbin/nologin`*.**
   b. **Create an FTP account. Run the *`/alidata/www/wwwroot`* command to specify the directory of user pwftp, or run the `useradd -d /alidata/www/wwwroot -s /sbin/nologin pwftp` command to specify an FTP account and its directory.**
   c. **Run the `passwd pwftp` command to change the account password.**
   d. **Run the `chown -R pwftp.pwftp /alidata/www/wwwroot` command to modify the permissions on the specified directory.**
3. **Configure VSFTP.**
   a. **Run the `vi /etc/vsftpd.conf` command to modify the VSFTP configuration file.**
   b. **Modify `anonymous_enable=YES` to `anonymous_enable=NO`.**
   c. **Uncomment the following configuration lines:**

      ```
      local_enable=YES write_enable=YES chroot_local_user=YES chroot_lis
      t_enable=YES chroot_list_file=/etc/vsftpd.chroot_list
      ```
   d. **Save the modifications and exit.**
   e. **Add the FTP account name to *`/etc/vsftpd.chroot_list`*. Save the modifications and exit.**
4. **Modify shell configurations. Use the `vi` editor to modify *`/etc/shells`*. If the file does not contain *`/usr/sbin/nologin`* or *`/sbin/nologin`* (depending on the current system configurations), add it to the file.**
5. **Restart VSFTP and perform a logon test.**
   a. **Run the `service vsftpd restart` command to restart VSFTP.**
   b. **Use the pwftp account to perform an FTP logon test. The directory is *`/alidata/www/wwwroot`*.**

2.12.4 Configure FTP through IIS in Windows Server 2003

This topic describes how to configure FTP through IIS in Windows Server 2003.

Procedure

1. **After you remotely connect to an ECS instance, right-click My Computer and choose Manage from the shortcut menu.**
2. **In the navigation pane, choose Local Users and Groups > Users to open the user list. Right-click the blank area of the user list and choose New User from the shortcut menu.**
3. **Enter your FTP username and password, and click Create**
4. **In the navigation pane, expand Internet Information Services (IIS) Management and delete the default FTP site. Right-click FTP Site and choose New > FTP Site (F).... Follow FTP Site Creation Wizard to create an FTP site.**
5. **Enter the home directory path, which is** `D:\websoft\www`
6. **Configure the permissions for the www folder of the new FTP site.**
   a. **Right-click the www folder and choose Properties from the shortcut menu.**
   b. **Click the Security tab. In the Group or user names area, select Users and set permissions.**
   c. **Click Advanced. Configure advanced security settings and click Apply.**
   d. **In the message that appears, click Yes.**
   e. **Wait until the process is completed.**
   f. **On the Security tab, add the permissions for the pwftp account.**
   g. **Click OK. Click Advanced. Configure advanced security settings. The method is the same as that for Users.**
   h. **After you configure the advanced security settings, click Apply, and then OK.**

2.12.5 Install and configure FTP in Windows Server 2008

This topic describes how to use an instance in Windows Server 2008 or later to build an FTP site.

Procedure

1. **After you start Windows Server, choose Start > Administrative Tools > Internet Information Services (IIS) Manager. Right-click the server name and choose Add FTP Site from the shortcut menu.**
2. **Enter the FTP site name and the specified path. Click Next.**

3. **Set** `IP Address` **to** `All Unassigned` **and** `SSL` **to** `No SSL`**.**
4. **Set** `Authentication` **to** `Basic`**,** `Authorization` **to** `All users`**, and** `Permissions` **to** `Read` **and** `Write`**.**
5. **Click Finish after you configure FTP settings. Then, you can use the administrator account and password to upload and download files through FTP.**

2.12.6 Install and configure IIS and FTP in Windows Server 2012

This topic describes how to install and configure IIS and FTP in Windows Server 2012.

Procedure

1. **In the lower-left corner of the Windows Server interface, click the Server Manager icon to start Server Manager.**
2. **Start the IIS manager.**
3. **Add an FTP site to the IIS manager.**
4. **Enter the FTP site name and specify the FTP path.**
5. **Set** `IP Address` **to** `All Unassigned` **and** `SSL` **to** `No SSL`
6. **Set** `Authentication` **to** `Basic`**,** `Authorization` **to** `All users`**, and** `Permissions` **to** `Read` **and** `Write`**.**
7. **After you configure FTP settings, use the default administrator account and password to perform a logon test. Then, you can upload and download files.**

3 Container Service

3.1 What is Container Service?

Container Service provides high-performance, enterprise-class management for scalable Kubernetes-based containerized applications throughout the application lifecycle.

Container Service simplifies the creation and scaling of container management clusters. It integrates Apsara Stack virtualization, storage, network, and security capabilities, providing the optimal environment to run Kubernetes-based containerized applications in the cloud. Alibaba Cloud is a Kubernetes certified service provider, with Container Service being among the first services to pass the Certified Kubernetes Conformance Program. Container Service provides profession al container support and services.

3.2 Planning and preparation

Before using Container Service, you must create resources such as VPCs, VSwitches, cloud disks, and OSS instances based on the cluster and application configuration requirements.

Before creating a Kubernetes cluster, you can make the following preparations:

- **(Optional) Create a VPC.**

  If you choose to create a cluster based on an existing VPC, create a VPC and a VSwitch first.
- **(Optional) Create a volume.**

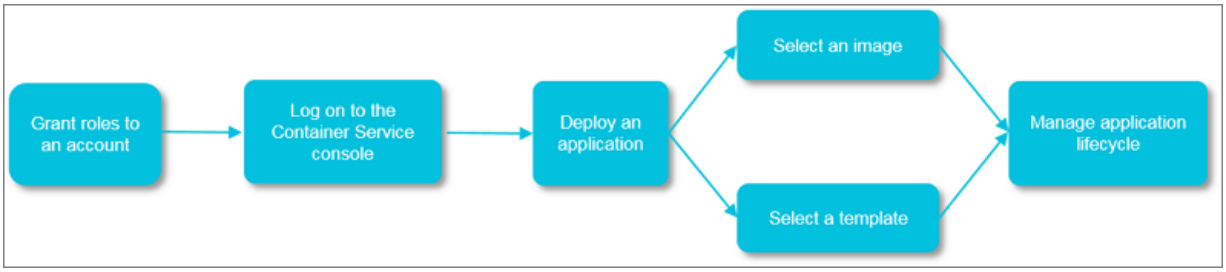
  If you want to create a volume for a stateful application based on network storage , create a cloud disk or OSS instance first.

3.3 Quick start

3.3.1 Procedure

You can use Container Service through the following procedure.

The following figure shows how to use Container Service.



Step 1: Authorize a role.

Grant permissions to the department that you want to use to log on to the Container Service console.

Step 2: Log on to the Container Service console.

Log on to the Container Service console. For more information, see *Log on to the Container Service console*.

Step 3: Create a cluster.

Select a network environment for the cluster. Set the number of nodes in the cluster and configure relevant information.

Step 4: Use an image or orchestration template to deploy an application.

Select an existing image or orchestration template, or create a new image or orchestration template.

If your application is composed of services supported by multiple images, use an orchestration template to create the application.

Step 5: Manage the application lifecycle.

3.3.2 Log on to the Container Service console

You can perform the following steps to log on to the Container Service console.

Prerequisites

- **Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://***IP address or domain name of the Apsara Stack console***/manage.**
- **We recommend that you use the Chrome browser.**

Procedure

1. **Open your browser.**

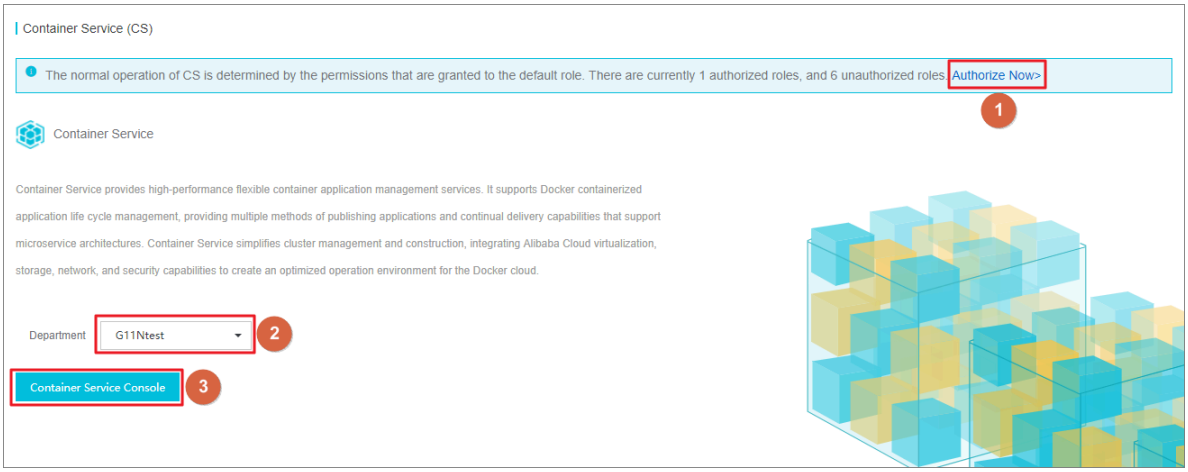
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click LOGIN to go to the Dashboard page.
5. In the top navigation bar, click  and choose Compute, Storage & Networking > Container Service.
6. On the Container Service page, grant permissions to the default role that is required to run Container Service.
   a) Click Authorized Now to authorize the current department.

 **Note:**

If you are logged on as an authorized department, skip this step.

b) In the select box below, select Department.

c) Click Docker to go to the Container Service console.



3.3.3 Create a Kubernetes cluster

To create a Kubernetes cluster, you must configure a series of cluster parameters. For more information, see *Table 3-1: Cluster parameters*.

Procedure

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, click Clusters. On the Clusters page, click Create Kubernetes Cluster in the upper-right corner.
3. Set the cluster parameters.

Table 3-1: Cluster parameters

Parameter	Description
Cluster Name	The name of the cluster to be created. The name must be 1 to 63 characters in length and can contain digits, letters, and hyphens (-). **Note:** The cluster name must be unique within the account.
Region	The region where the cluster is to be deployed.
Zone	The zone where the cluster is located.

Parameter	Description
VPC	You can select Auto Create to create a VPC automatically when the Kubernetes cluster is created, or select Use Existing to use an existing VPC. If you select Use Existing, select the required VPC and VSwitch from the VPC list. · If you select Auto Create, the system automatically creates a NAT Gateway for your VPC. · If you select Use Existing and the selected VPC already has a NAT Gateway, Container Service will use the existing NAT Gateway. If the selected VPC has no NAT Gateways, you can set a NAT Gateway on the VPC page.  Note: If automatic NAT Gateway creation is disabled in the system, you must manually configure the NAT Gateway or SNAT to ensure that the VPC can connect to the public network properly. Otherwise, instances in the VPC cannot connect to the public network, which will lead to cluster creation failures. You must specify Pod Network CIDR and `Service CIDR`. The two CIDRs cannot be the same as CIDR blocks used by the VPC or existing Kubernetes clusters in the VPC. The CIDRs cannot be modified after cluster creation. The service CIDR block cannot be the same as the pod CIDR block.

Parameter	Description
Node Type	Kubernetes clusters only support Pay-As-You-Go nodes.
Master Configuration	Select an instance type and a system disk. · Instance Type: For more information about instance types, see the Instance type family section in ECS *User Guide*. · System Disk: Available options are SSD Disk and Ultra Disk.
Worker Instance	You can create Worker instances or add existing ones.
Worker Configuration	If you choose to add an instance in the Worker Instance section, the configuration is as follows: · Instance Type: For more information about instance types, see the Instance type family section in ECS *User Guide*. · System Disk: Available options are SSD Disk and Ultra Disk. · Attach Data Disk: Available options are SSD Disk, Ultra Disk, and Basic Disk.
Logon Password	Set the password used to log on to the node.
Confirm Password	Confirm the password used to log on to the node.
Pod Network CIDR and Service CIDR (optional)	For more information about the specific plan, see *Plan Kubernetes CIDR blocks in a VPC*. **Note:** These options are available when you select Use Existing.

Parameter	Description
Configure SNAT	If you select Auto Create, you must configure SNAT. If you select Use Existing, you can choose whether to automatically configure SNAT. If you choose not to automatically configure SNAT, you must manually configure the NAT Gateway or SNAT.
SSH Logon	· If you choose to enable SSH access through the public network, you can access a cluster through SSH. · If you choose to disable SSH access through the public network, the cluster cannot be accessed through SSH or kubectl. You can manually enable SSH access. For more information, see *Access a Kubernetes cluster through SSH*.
RDS Whitelist (optional)	Add the IP addresses of the ECS instances to the RDS instance whitelist. **Note:** This option is available when you select Use Existing.
Advanced Options	· Pod Number for Node: indicates the maximum number of pods that can be run on a single node. · Cluster CA: allows you to specify whether to use a custom cluster CA.

4. Click Create Cluster in the upper-right corner of the cluster configuration page.
5. On the Confirm Cluster Configuration page, check all parameter settings and click Confirm.

Result

After the cluster is created, you can view it on the Clusters page of the Container Service console.

3.3.4 Create an application from an orchestration template

Container Service provides orchestration templates that you can use to create applications quickly. You can also modify the templates based on YAML syntax to customize your applications.

Prerequisites

You have created a Kubernetes cluster. For more information, see *Create a Kubernetes cluster*.

Context

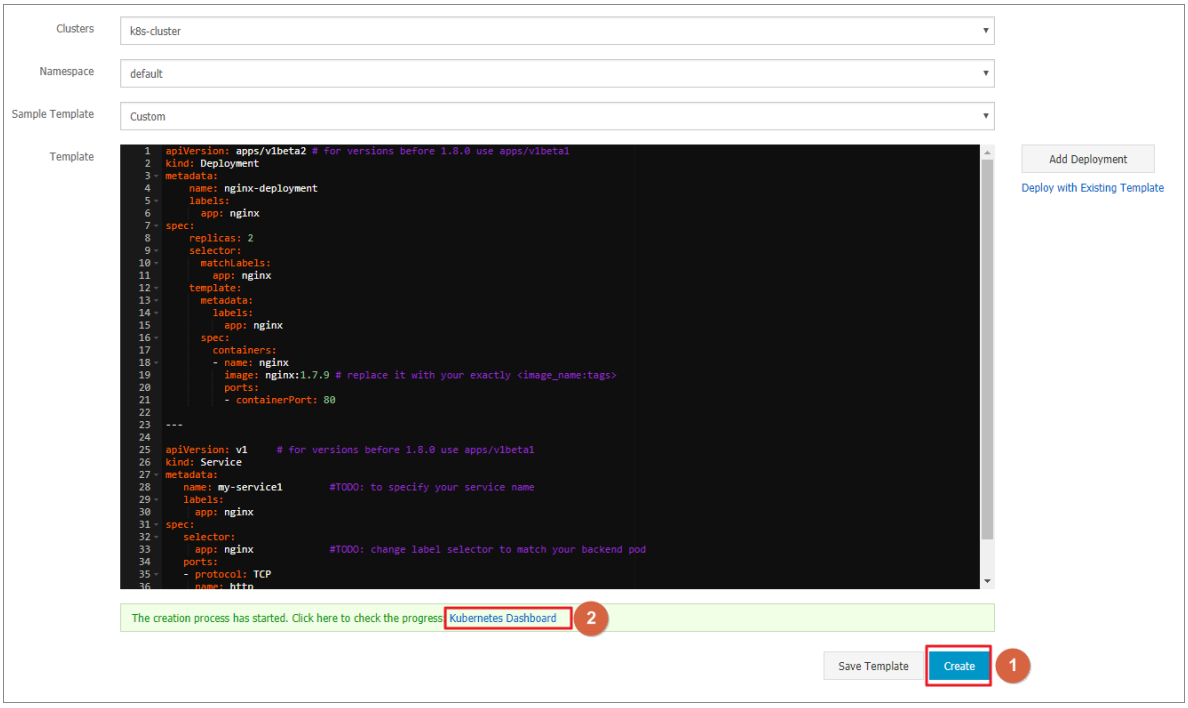
The following example demonstrates how to create an Nginx application that consists of a Deployment and a Service. The Deployment creates a Pod and the Service is then associated with the Pod.

Procedure

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Applications > Deployments to go to the Deployments page.**
3. **In the upper-right corner, click Create from Template.**
4. **Set the parameters and click Create.**
   - **Cluster: Select the cluster where the resource objects are to be deployed.**
   - **Namespace: Select the namespace that the resource objects belong to. The default namespace is default. Except for underlying computing resources such as nodes and persistent volumes, most resource objects must be divided into namespaces.**
   - **Sample Template: Container Service provides YAML templates of various resource types to help you deploy resource objects quickly. You can create a**

template based on YAML syntax to describe the resource types that you want to define.

- Add Deployment: You can quickly define a YAML template.
- Use Existing Template: You can import an existing template to the configuration page.



The following is a sample orchestration of an Nginx application. The orchestration is based on an orchestration template provided by Container Service. You can use this orchestration template to create a Deployment for an Nginx application quickly.

Note:

Container Service supports the YAML syntax and supports using the `---` symbol to separate resource objects. This allows you to create multiple resource objects in a single template.

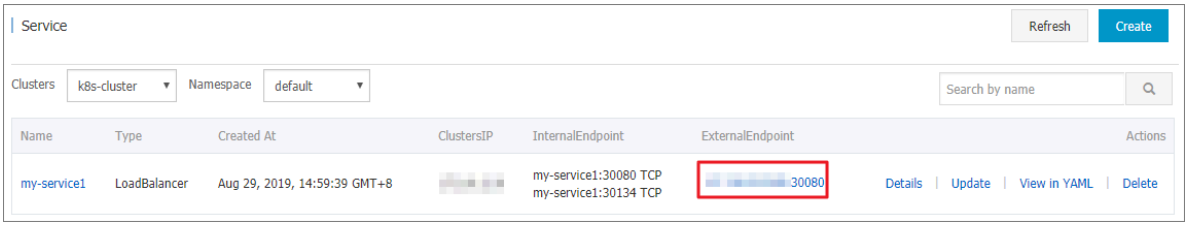
```
apiVersion: apps/v1beta2 # for versions before 1.8.0 use apps/
v1beta1
kind: Deployment
metadata:
    name: nginx-deployment
    labels:
      app: nginx
spec:
    replicas: 2
    selector:
      matchLabels:
        app: nginx
```

```
    template:
      metadata:
        labels:
          app: nginx
      spec:
        containers:
        - name: nginx
          image: nginx:1.7.9 # replace it with your exactly <
image_name:tags>
          ports:
          - containerPort: 80

---

apiVersion: v1      # for versions before 1.8.0 use apps/v1beta1
kind: Service
metadata:
   name: my-service1        #TODO: to specify your service name
   labels:
     app: nginx
spec:
   selector:
     app: nginx             #TODO: change label selector to match
your backend pod
   ports:
   - protocol: TCP
     name: http
     port: 30080            #TODO: choose an unique port on each
node to avoid port conflict
     targetPort: 80
   type: LoadBalancer        ##In the example, the type is changed
from Nodeport to LoadBalancer.
```

5. **Click Create. A message appears indicating the deployment status. You can click Kubernetes Dashboard to check the deployment progress in the dashboard.**
6. **On the Kubernetes dashboard, you can see that a Service named my-service1 is deployed and its external endpoint is displayed. Click the address under External Endpoint.**



7. **You can visit the Nginx welcome page in the browser.**

What's next

You can also choose Ingresses and Load Balancing > Services in the left-side navigation pane to view the Nginx service.

3.4 Kubernetes clusters

3.4.1 Clusters

3.4.1.1 Create a Kubernetes cluster

To create a Kubernetes cluster, you must configure a series of cluster parameters. For more information, see *Table 3-2: Cluster parameters*.

Procedure

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, click Clusters. On the Clusters page, click Create Kubernetes Cluster in the upper-right corner.**
3. **Set the cluster parameters.**

Table 3-2: Cluster parameters

Parameter	Description
Cluster Name	The name of the cluster to be created. The name must be 1 to 63 characters in length and can contain digits, letters, and hyphens (-). **Note:** The cluster name must be unique within the account.
Region	The region where the cluster is to be deployed.
Zone	The zone where the cluster is located.

Parameter	Description
VPC	You can select Auto Create to create a VPC automatically when the Kubernetes cluster is created, or select Use Existing to use an existing VPC. If you select Use Existing, select the required VPC and VSwitch from the VPC list. · If you select Auto Create, the system automatically creates a NAT Gateway for your VPC. · If you select Use Existing and the selected VPC already has a NAT Gateway, Container Service will use the existing NAT Gateway. If the selected VPC has no NAT Gateways, you can set a NAT Gateway on the VPC page.  Note: If automatic NAT Gateway creation is disabled in the system, you must manually configure the NAT Gateway or SNAT to ensure that the VPC can connect to the public network properly. Otherwise, instances in the VPC cannot connect to the public network, which will lead to cluster creation failures. You must specify Pod Network CIDR and `Service CIDR`. The two CIDRs cannot be the same as CIDR blocks used by the VPC or existing Kubernetes clusters in the VPC. The CIDRs cannot be modified after cluster creation. The service CIDR block cannot be the same as the pod CIDR block.

Parameter	Description
Node Type	Kubernetes clusters only support Pay-As-You-Go nodes.
Master Configuration	Select an instance type and a system disk. · Instance Type: For more information about instance types, see the Instance type family section in ECS *User Guide*. · System Disk: Available options are SSD Disk and Ultra Disk.
Worker Instance	You can create Worker instances or add existing ones.
Worker Configuration	If you choose to add an instance in the Worker Instance section, the configuration is as follows: · Instance Type: For more information about instance types, see the Instance type family section in ECS *User Guide*. · System Disk: Available options are SSD Disk and Ultra Disk. · Attach Data Disk: Available options are SSD Disk, Ultra Disk, and Basic Disk.
Logon Password	Set the password used to log on to the node.
Confirm Password	Confirm the password used to log on to the node.
Pod Network CIDR and Service CIDR (optional)	For more information about the specific plan, see *Plan Kubernetes CIDR blocks in a VPC*. **Note:** These options are available when you select Use Existing.

Parameter	Description
Configure SNAT	If you select Auto Create, you must configure SNAT. If you select Use Existing, you can choose whether to automatically configure SNAT. If you choose not to automatically configure SNAT, you must manually configure the NAT Gateway or SNAT.
SSH Logon	· If you choose to enable SSH access through the public network, you can access a cluster through SSH. · If you choose to disable SSH access through the public network, the cluster cannot be accessed through SSH or kubectl. You can manually enable SSH access. For more information, see *Access a Kubernetes cluster through SSH*.
RDS Whitelist (optional)	Add the IP addresses of the ECS instances to the RDS instance whitelist. **Note:** This option is available when you select Use Existing.
Advanced Options	· Pod Number for Node: indicates the maximum number of pods that can be run on a single node. · Cluster CA: allows you to specify whether to use a custom cluster CA.

4. Click Create Cluster in the upper-right corner of the cluster configuration page.
5. On the Confirm Cluster Configuration page, check all parameter settings and click Confirm.

Result

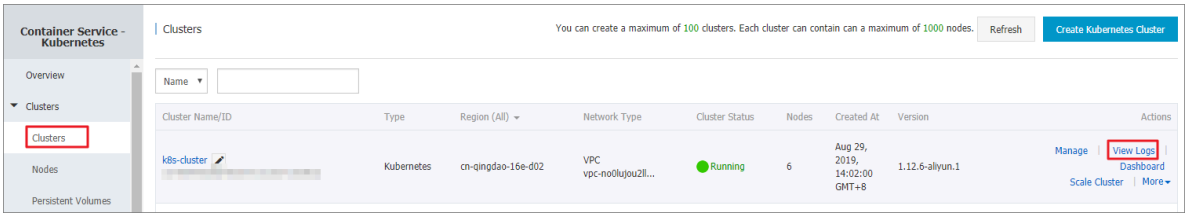
After the cluster is created, you can view it on the Clusters page of the Container Service console.

3.4.1.2 View cluster logs

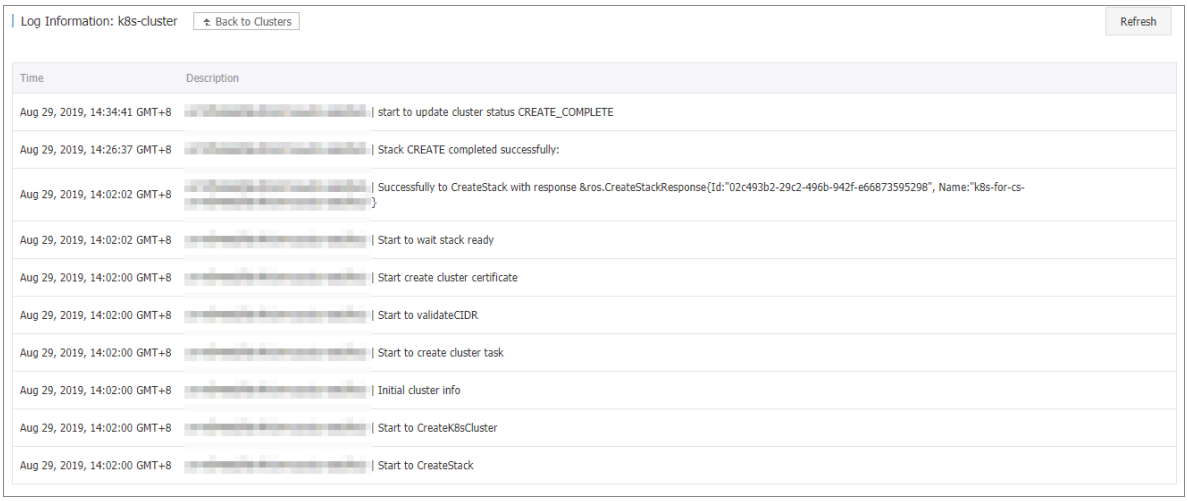
You can view cluster logs through the Container Service simple log service.

Procedure

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, click Clusters. The Clusters page appears.
3. Click View Logs in the Actions column corresponding to a cluster.



View the cluster operation information.



3.4.1.3 Connect to a Kubernetes cluster through kubectl

To connect to a Kubernetes cluster from a client computer, use the Kubernetes command line client *kubectl*.

Procedure

1. Go to the *Kubernetes version page* to download the latest kubectl client.
2. Install and set the kubectl client.

   For more information, see *Install and set kubectl*.
3. Configure the cluster credentials.

   You can use the `scp` command to copy the master node configurations from `/etc/kubernetes/kube.conf` in the master VM of the Kubernetes cluster to `$HOME/.`

`kube/config` on the local computer where the expected `kubectl` credentials are located.

```
mkdir $HOME/.kube
scp root@<master-public-ip>:/etc/kubernetes/kube.conf $HOME/.kube/config
```

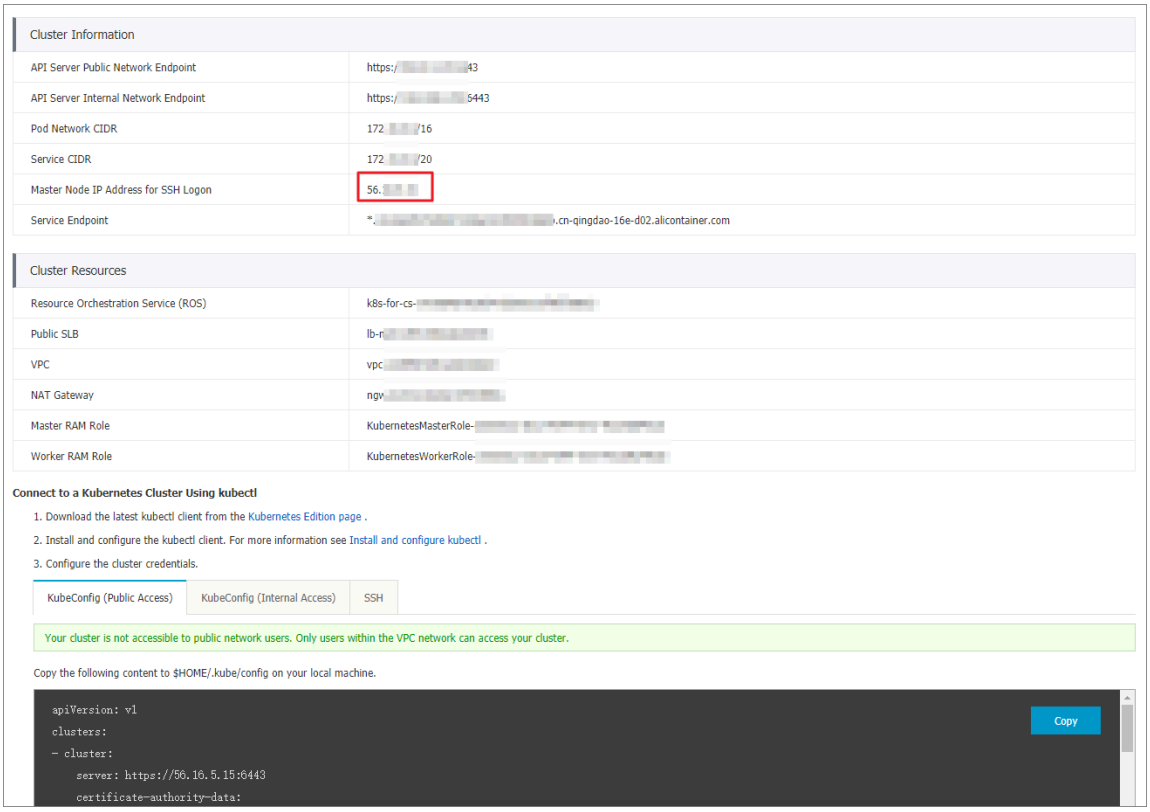
You can then view the `master-public-ip` parameter of the cluster on the cluster information page.

a) *Log on to the Container Service console*.

b) In the left-side navigation pane, click Clusters. The Clusters page appears.

c) Click Management in the Actions column corresponding to a cluster.

You can view the cluster access URL in the Connection Information section.



3.4.1.4 Connect to a master node through SSH

The master node can be accessed through the IP address used to enable SSH access to the cluster.

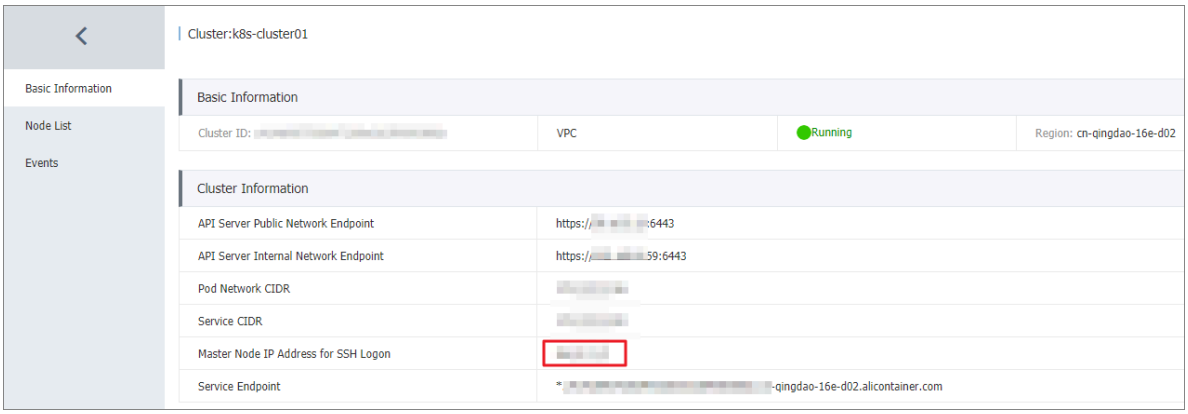
Prerequisites

- You have created a Kubernetes cluster with the Enable SSH access for Internet check box selected. For more information, see *Create a Kubernetes cluster*.

- You must be in a network environment that can communicate with the Kubernetes cluster network.

Procedure

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, click Clusters. On the Clusters page, click Management in the Actions column corresponding to a cluster.
3. On the cluster details page, view the Master node SSH IP address.



4. In the network environment that can communicate with the Kubernetes cluster network, connect to the master node through SSH.
   - If you have configured a leased line, you can use tools such as Putty to initiate an SSH connection to the Kubernetes cluster network through the Internet.
   - If you have a server that can communicate with the cluster VPC, such as an ECS instance, you can run the following command:

```
ssh    root@ssh_ip                          #ssh_ip is the Master node
 SSH IP address.
```

3.4.1.5 Cluster scaling

You can use add or remove worker nodes to or from Kubernetes clusters as needed in the console.

Context

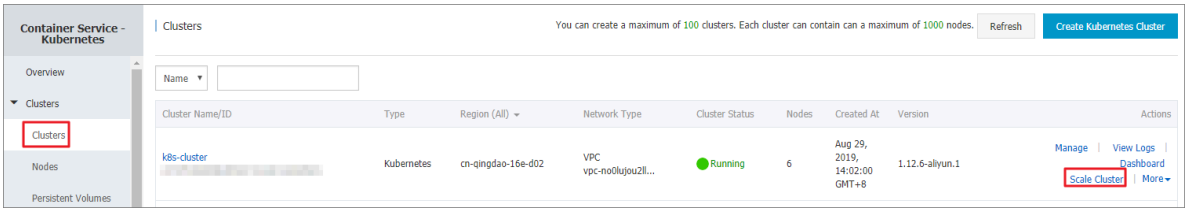
- Currently, you can only scale clusters manually. Automatic scaling is not supported.
- Currently, you cannot add or remove master nodes to or from clusters.
- Scaling in a cluster only removes worker nodes that were added when you created or scaled out the cluster. These worker nodes cannot be removed by

using the `kubectl delete` command or through the console. Worker nodes that were added to the cluster through the Add Existing Node option cannot be removed when you scale in the cluster.

- Nodes are removed from the cluster in the reverse order that they were added. The most recently added node is reclaimed first.
- The cluster must have at least two nodes that were not manually added.

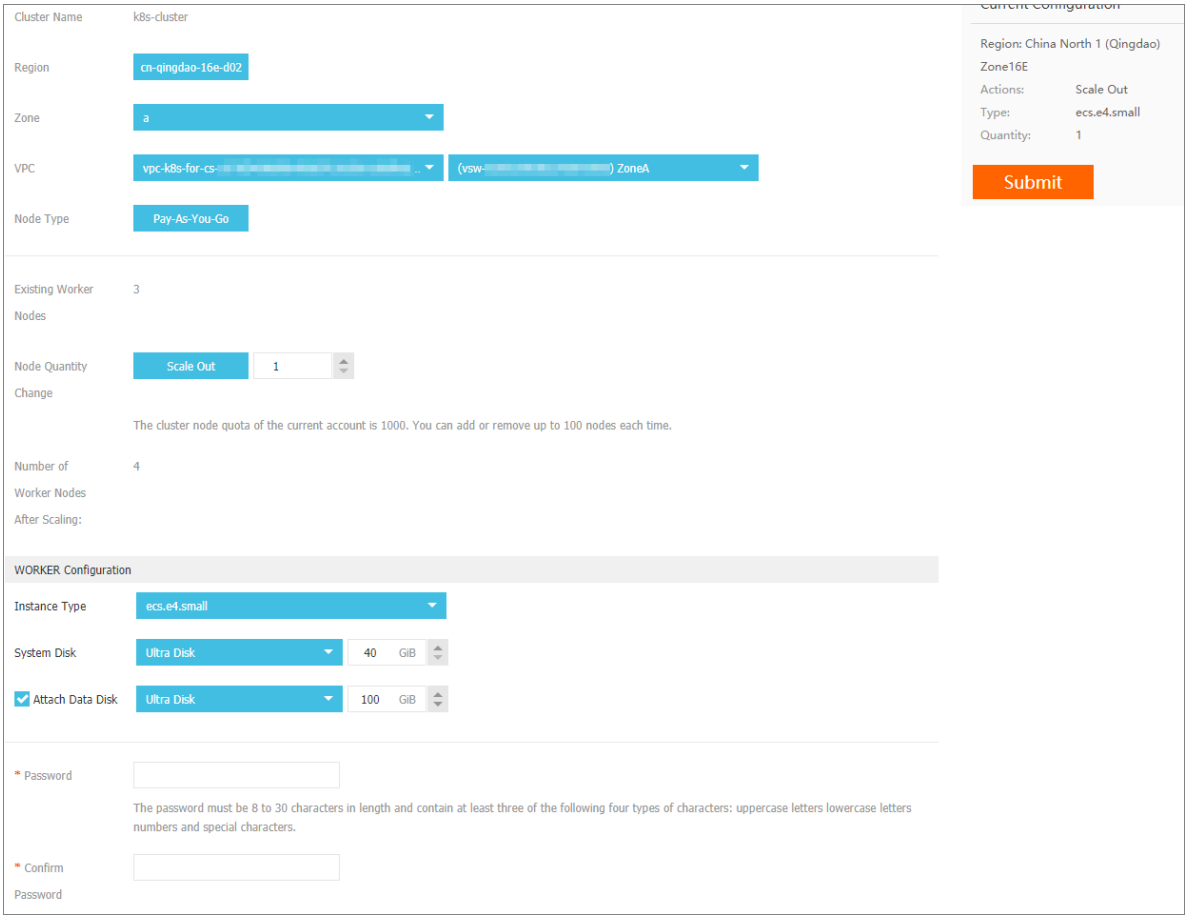
Procedure

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Clusters > Clusters to go to the Clusters page.
3. Select the cluster and click Scale Cluster in the Actions column.



4. Select Scale Out or Scale In and specify the number of worker nodes.

   This example describes how to increase the number of worker nodes in the cluster from 6 to 8.



5. Enter the logon password of the node.

   > **Note:**
   > Make sure that the password is the same as the one you entered when you created the cluster. You will need the password to log on to the ECS instance to copy configuration information in the upgrade process.

6. Click Submit.

What's next

In the left-side navigation pane, choose Clusters > Nodes to go to the Nodes page. The number of worker nodes is now changed to 8.

3.4.1.6 Update certificates

This topic describes how to update certificates for your clusters in the console.

Prerequisites

You have created a Kubernetes cluster and its certificate is about to expire soon.

Procedure

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Clusters > Clusters to go to the Clusters page.
3. Select the cluster and click Update Certificate to go to the Update Certificate page.

   Note:
   The Update Certificate option will be available when your certificate is going to expire in about two months.
4. Click Update and the Confirm page appears.
5. Click OK.

Result

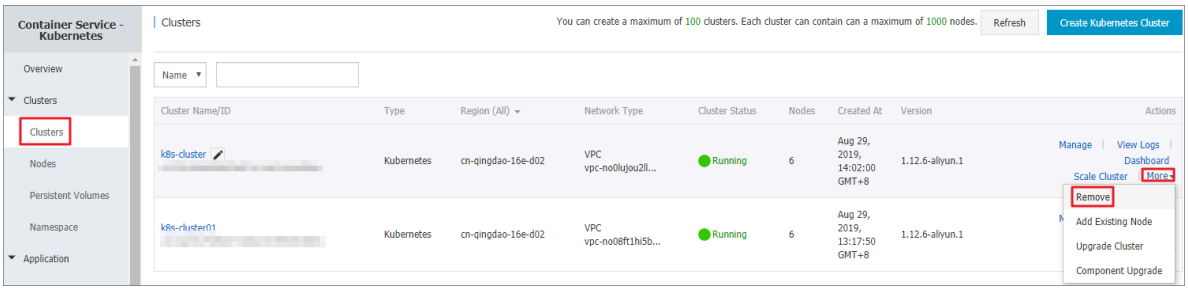
- On the Update Certificate page, a message appears indicating that the certificate is updated.
- On the Clusters page, the Update Certificate option has disappeared.

3.4.1.7 Delete clusters

You can delete clusters in the Container Service console.

Procedure

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Clusters > Clusters to go to the Clusters page.
3. Select a cluster and click Delete in the Actions column.



What's next

Failed to delete the cluster

ROS has no permission to delete resources that were manually added under ROS-created resources. For example, if you manually add a VSwitch under a ROS-created VPC instance, ROS cannot delete the VPC instance and therefore the cluster cannot be deleted.

Container Service allows you to force delete clusters. This function enables you to force delete a cluster and ROS resource stack if your first attempt to delete the cluster fails. However, you still need to manually release the resources that were manually added in the first place.

An error message appears when the attempt to delete a cluster fails.

Select the cluster that you failed to delete and click Delete in the Actions column. In the dialog box that appears, you can view information about the resources that were manually added. Click Force Delete and click OK to delete the cluster and ROS resource stack.

Note:

You must manually release the resources that were manually added in the first place.

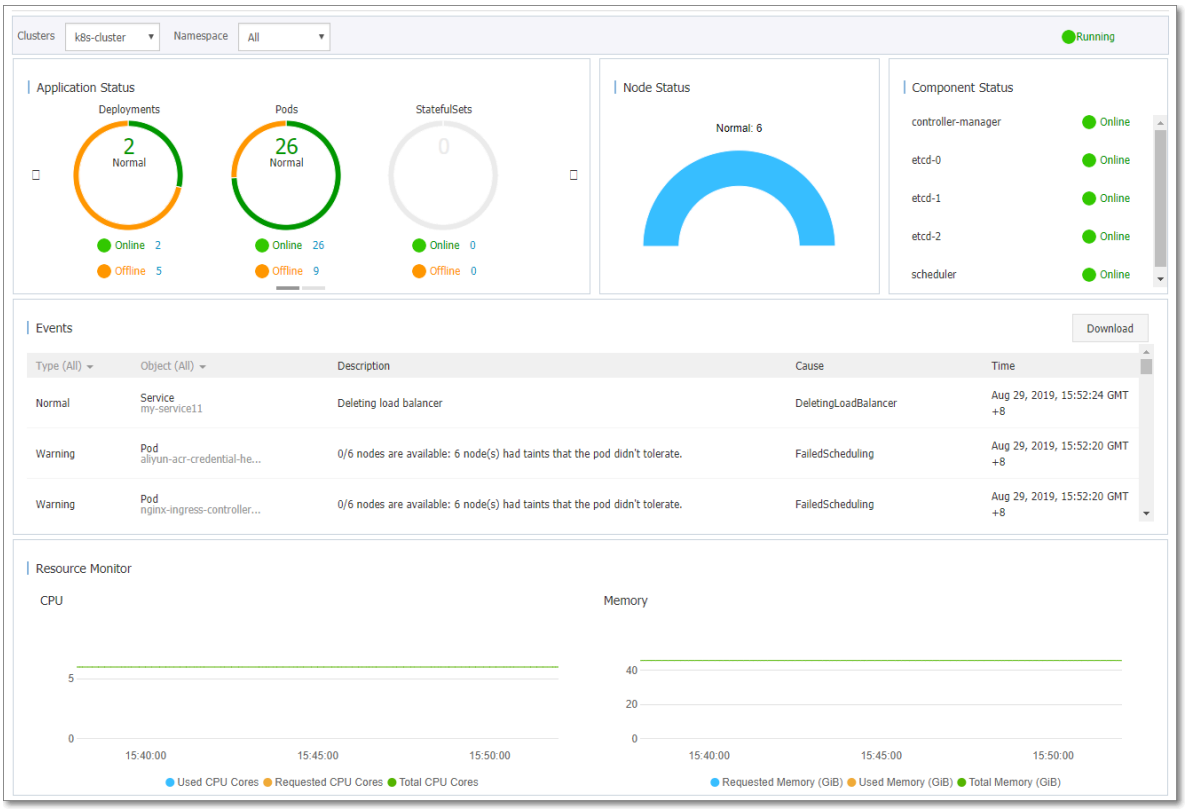
3.4.1.8 View cluster overview

On the Overview page of Apsara Stack Container Service Kubernetes clusters, you can view resource monitoring charts as well as the statuses of applications and components to monitor the health statuses of clusters.

Procedure

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, click Overview. The Kubernetes cluster overview page appears.

3. Select a cluster and a namespace from the Cluster and Namespace drop-down lists. You can view resource monitoring charts as well as the statuses of applications and components.
   - Application status: indicates the status of deployments, pods, and replica sets that are currently running. Green indicates the application is in the normal status, and orange indicates an exception.
   - Node status: indicates the node status of the current cluster.
   - Component status: indicates the status of Kubernetes cluster components that are typically deployed in the kube-system namespace, including core components such as the scheduler, controller-manager, and etcd.
   - Resource monitor: provides the CPU and memory monitoring charts. CPU is measured in cores and is accurate to three decimal places. The minimum unit is millicores, or one thousandth of one core. Memory is measured in GB and is accurate to three decimal places. For more information, see *Meaning of CPU* and *Meaning of memory*.
   - Events: provides event information of the cluster, such as warnings and error events. If the cluster is in the normal status, no data is displayed.



3.4.2 Nodes

3.4.2.1 Add existing nodes

You can add existing ECS instances to clusters. Currently, you can only add worker nodes to clusters.

Prerequisites

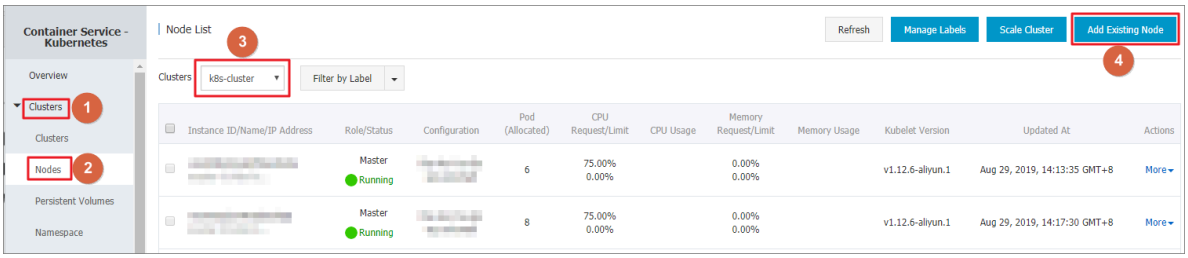
- **If you have not created a cluster before, you need to *Create a Kubernetes cluster*.**
- **You must have created an ECS instance and make sure that the region, zone, department, project, security group, VPC network, and operating system settings of the ECS instance are the same as those of the cluster.**

Context

- **By default, a cluster can contain up to 40 nodes. To add more nodes, submit a ticket.**
- **The ECS instance must be in the same region and VPC network as the cluster.**
- **The ECS instance must belong to the same account as the cluster.**
- **The ECS instance must be running the CentOS operating system.**

Procedure

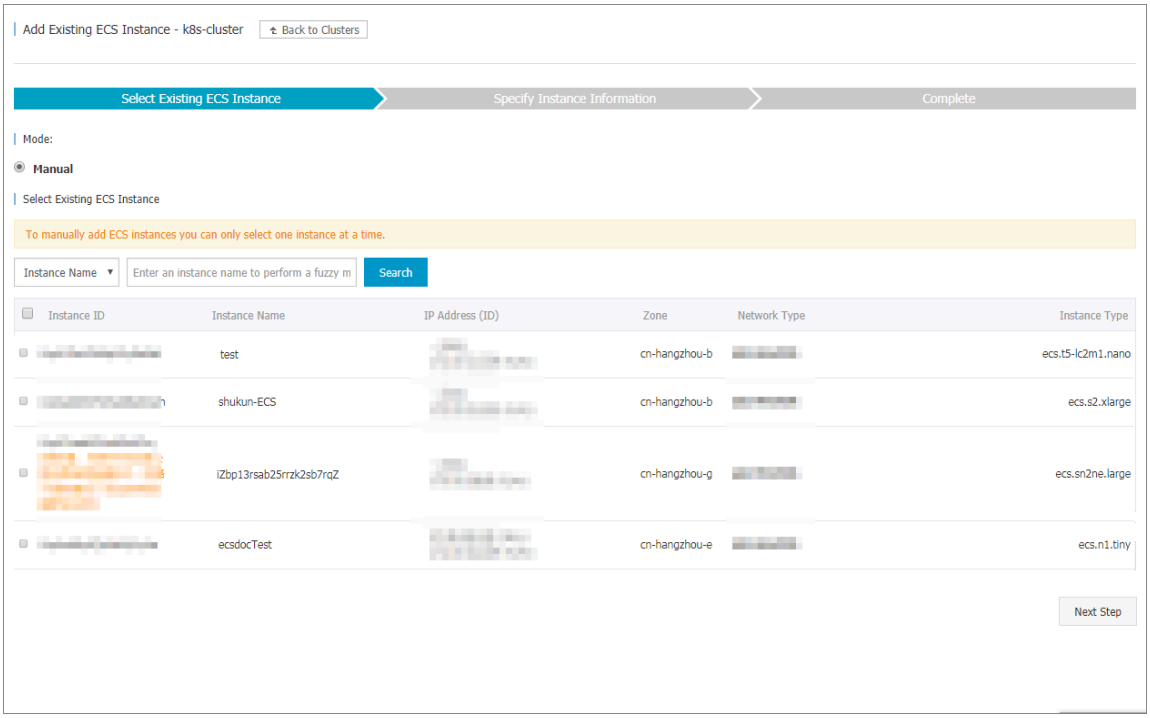
1. ***Log on to the Container Service console*.**
2. **In the left-side navigation pane, choose Clusters > Nodes to go to the Clusters page.**
3. **Select a cluster, and click Add Existing Node in the upper-right corner.**



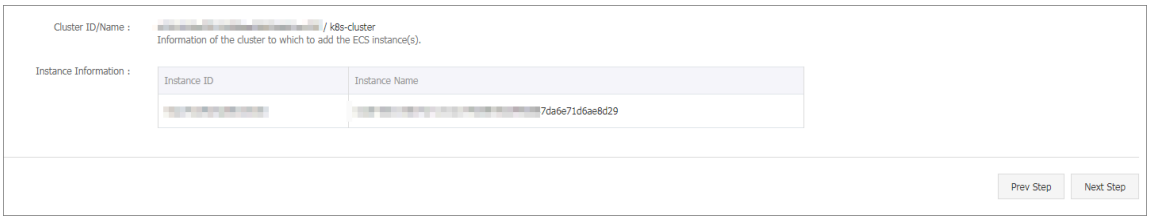
4. On the Add Existing Node page, you can manually add existing ECS instances.

   To manually add an ECS instance, you need to obtain the installation command and log on to the ECS instance to run the command. You can only add one ECS instance at a time.

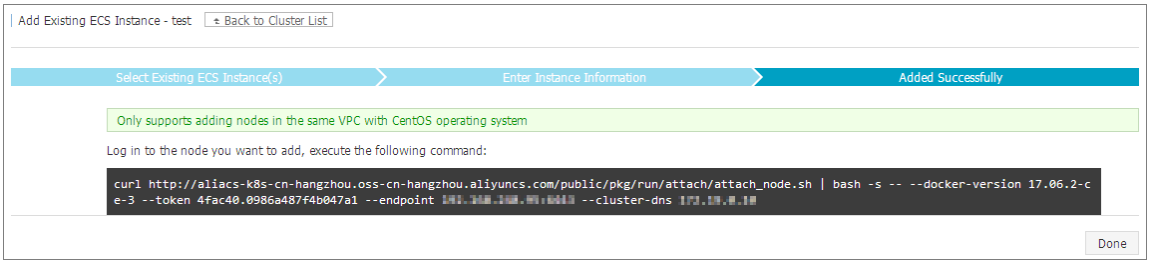
   a) Select the ECS instance that you want to add and click Next Step. You can add one ECS instance at a time.



   b) Confirm the instance information and click Next Step.



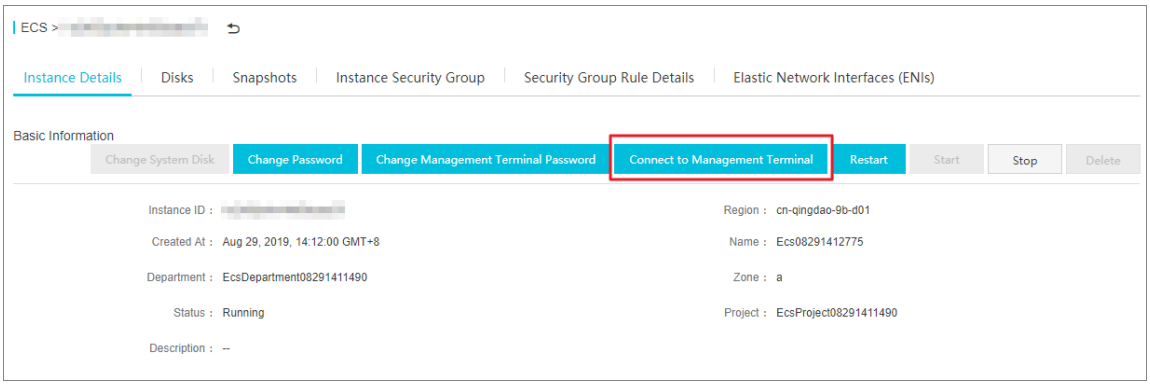
   c) Copy the command.



   d) Click Complete.

e) Go to the Apsara Stack console and click Compute, Storage & Networking > ECS. In the top navigation bar, click Instance. Select the department and region of the cluster, and find the ECS instance that you want to add.

f) Click the instance name to go to the details page. Click Log on to VNC. The Enter VNC Password dialog box appears. Enter the VNC password and then click OK. Enter the copied command and click OK to run the script.



g) After the script is successfully executed, the ECS instance is added to the cluster. You can click the cluster ID on the Clusters page to view the nodes in the cluster and check whether the ECS instance has been added to the cluster.

3.4.2.2 View nodes

You can view the nodes in a Kubernetes cluster though commands, the Container Service console, or the Kubernetes dashboard.

Run a command to view nodes

Note:

Before running the command to view the nodes in the Kubernetes cluster, see *Connect to a Kubernetes cluster through kubectl*.

Connect to the Kubernetes cluster through kubectl and run the following command to view the nodes in the cluster:

```
kubectl get nodes
```

Sample command output:

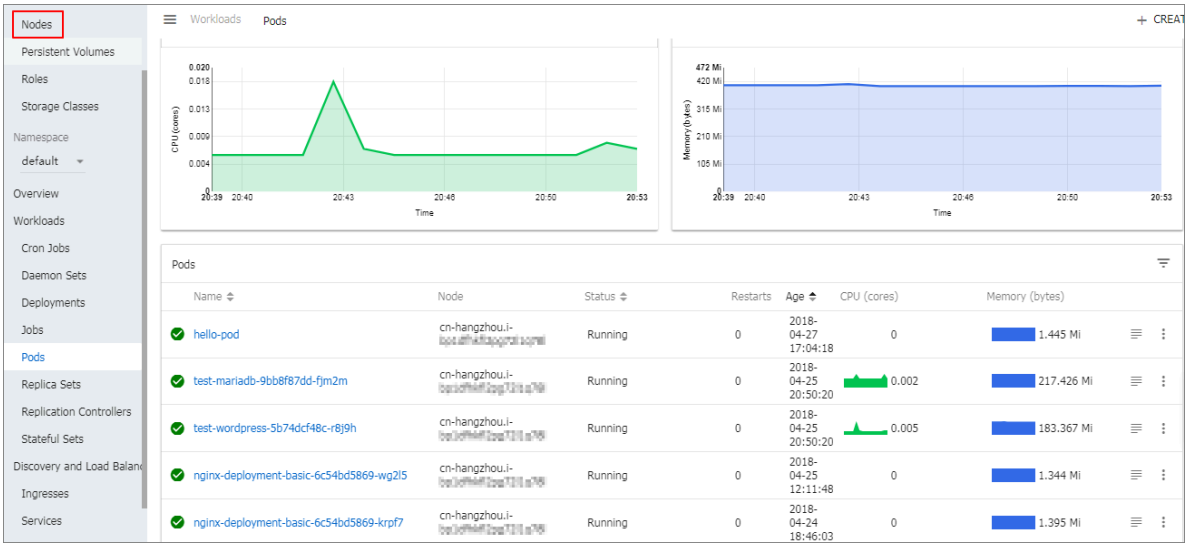
```
$ kubectl get nodes
    NAME STATUS AGE VERSION
    iz2ze2n6ep53tch701yh9zz Ready 19m v1.6.1-2+ed9e3d33a07093
    iz2zeafr762wibijx39e5az Ready 7m v1.6.1-2+ed9e3d33a07093
    iz2zeafr762wibijx39e5bz Ready 7m v1.6.1-2+ed9e3d33a07093
    iz2zef4dnn9nos8elyr32kz Ready 14m v1.6.1-2+ed9e3d33a07093
    iz2zeitvvo8enoreufstkmz Ready 11m v1.6.1-2+ed9e3d33a07093
```

Use the Container Service console to view the nodes

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Clusters > Nodes. The Nodes page appears.**
3. **Select a cluster from the Cluster drop-down list. You can then view the nodes in the cluster.**

Use the Kubernetes dashboard to view nodes

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, click Clusters. The Clusters page appears.**
3. **Locate a cluster and click Console in the Actions column to go to the Kubernetes dashboard.**
4. **In the left-side navigation pane of the Kubernetes dashboard, click Nodes. You can then view the nodes in the cluster.**



3.4.2.3 Manage node labels

In the Container Service console, you can perform many operations to manage node labels, such as adding multiple node labels at a time, filtering nodes by label, and deleting node labels.

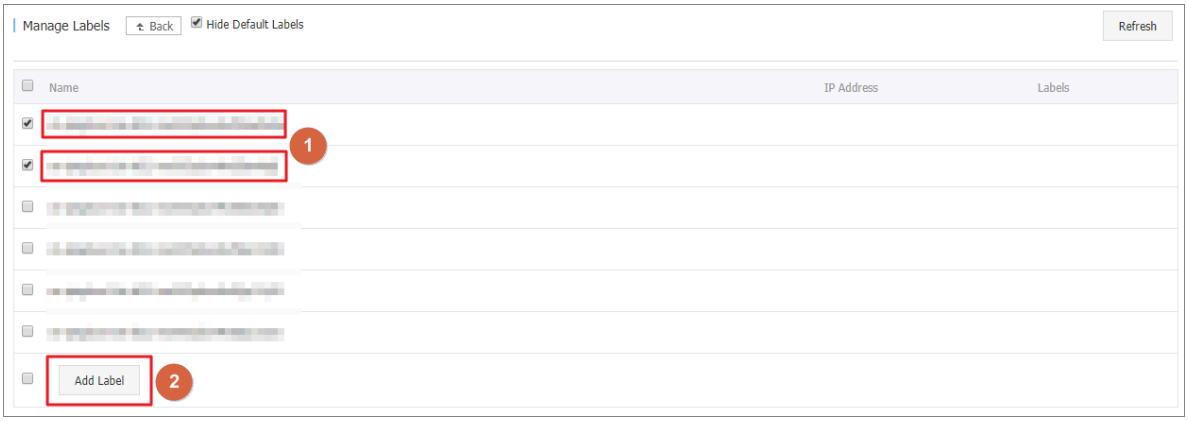
For more information about how to use node labels to schedule pods to specified nodes, see *Set node scheduling*.

Prerequisites

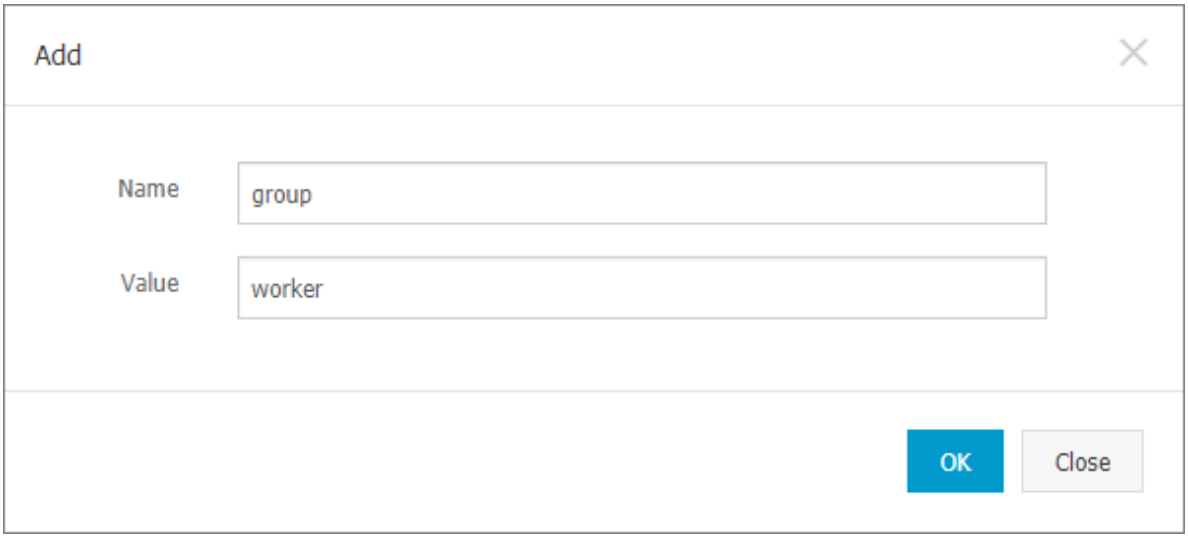
You have created a Kubernetes cluster. For more information, see *Create a Kubernetes cluster*.

Add multiple node labels at a time

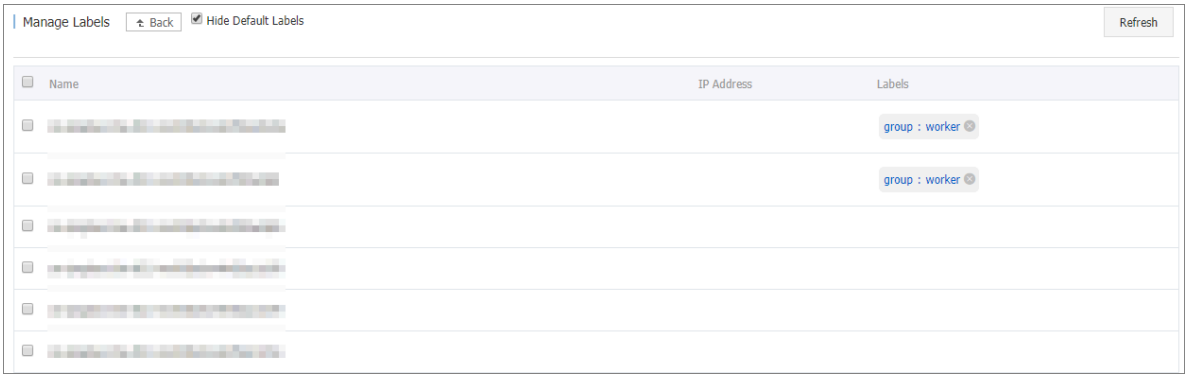
1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Clusters > Nodes. The Nodes page appears.**
3. **Select a cluster from the Cluster drop-down list and then click Label Management in the upper-right corner.**
4. **Select one or more nodes and then click Add Label.**



5. **In the dialog box that appears, enter the label name and value and then click OK.**



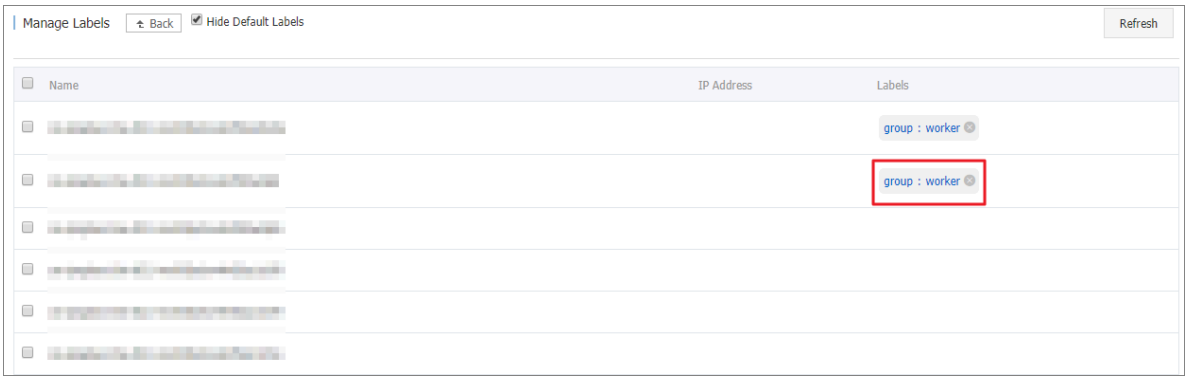
Nodes with the same label are displayed on the Label Management page.



Delete a node label

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Clusters > Nodes. The Nodes page appears.**
3. **Select a cluster from the Cluster drop-down list and then click Label Management in the upper-right corner.**

4. Click the x icon next to a node label such as `group:worker`.

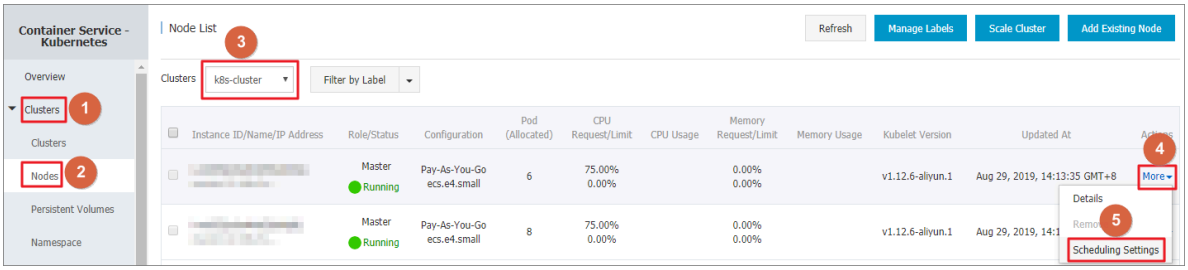


In the message that appears, click OK. You can verify that the node label is removed.

3.4.2.4 Set node scheduling

Procedure

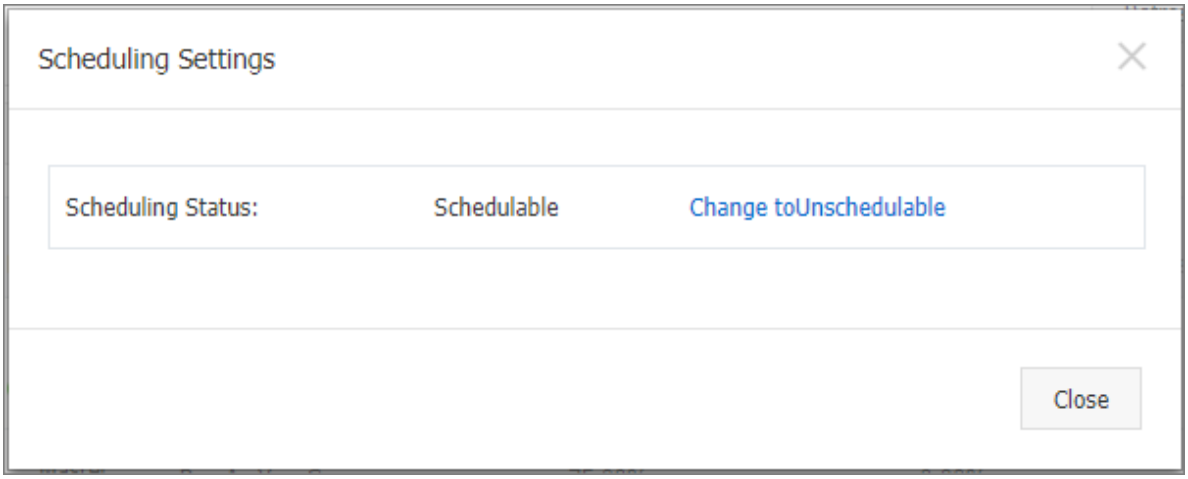
1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Clusters > Nodes. The Nodes page appears.
3. Select a cluster from the Cluster drop-down list. Then, click Scheduling Settings in the Actions column corresponding to a node.



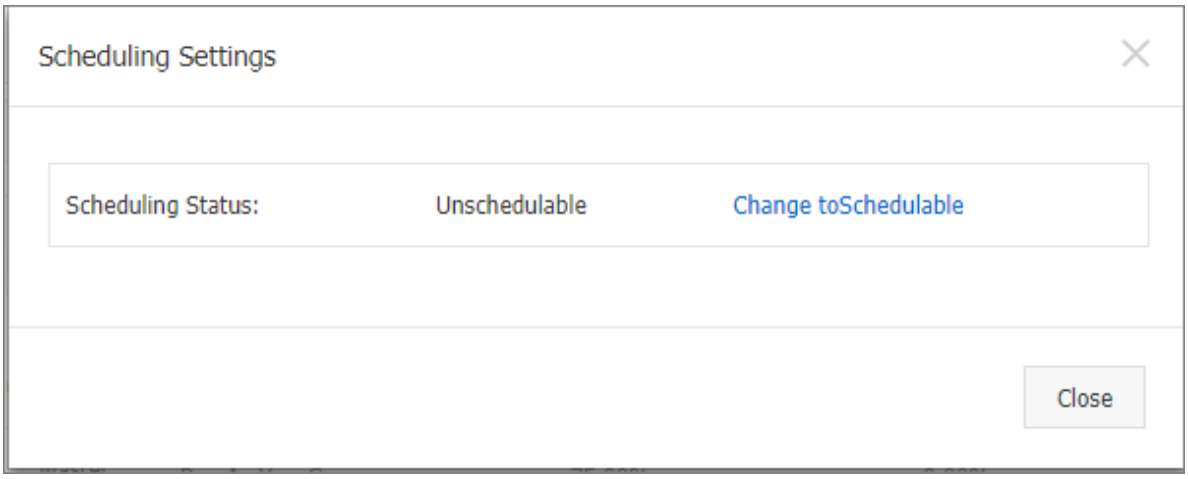
4. In the dialog box that appears, set the scheduling parameters. In this example, click Change to Unschedulable to set the node scheduling status to unschedulable.

Note:

The scheduling status of the current node is displayed in the Scheduling Settings dialog box. This status is schedulable by default. You can change the scheduling status of the node as needed.



After you change the status in the dialog box, the scheduling status of the node will change.



What's next

Later, when you deploy applications, pods will not be scheduled to the node.

3.4.2.5 View the resource requests and limits on nodes

The Container Service console allows you to view resource usage of each node in a Kubernetes cluster.

Prerequisites

You have created a Kubernetes cluster. For more information, see *Create a Kubernetes cluster*.

Procedure

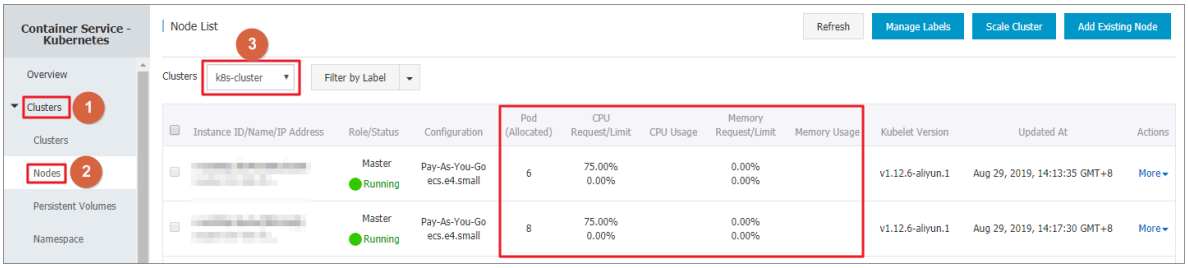
1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Clusters > Nodes. The Nodes page appears.

   You can view the CPU utilization and memory usage of each node, namely, the request and limit, which are calculated based on the following formulas:

   - CPU request (%) = Sum of CPU request values for all pods on the current node/ Total CPU capacity of the node.
   - CPU limit (%) = Sum of CPU limit values for all pods on the current node/Total CPU capacity of the node.
   - Memory request (%) = Sum of memory request values for all pods on the current node/Total memory capacity of the node.
   - Memory limit (%) = Sum of memory limit values for all pods on the current node/Total memory capacity of the node.

   Note:

   - You can allocate loads to nodes based on their resource usage. For more information, see *Set node scheduling*.
   - When either the resource request or usage rate of a node reaches 100%, new pods cannot be scheduled to that node.



3.4.3 Storage

3.4.3.1 Overview

In the Container Service console, you can create volumes of other Apsara Stack services, enabling you to create stateful applications and use Apsara Stack disks and OSS to implement persistent storage.

Both static and dynamic volumes are supported. The following table shows how static and dynamic volumes are supported.

Apsara Stack storage	Static volume	Dynamic volume
Apsara Stack disk	You can use a static disk volume through either of the following methods: · Use a volume directly · Use a volume through a PV and PVC	Supported
Apsara Stack NAS	You can use a static NAS volume through either of the following methods: · Use a volume through the FlexVolume plug-in - Use a volume directly - Use a volume through a PV or PVC · Use a volume through the Kubernetes NFS driver	Supported
Apsara Stack OSS	You can use a static OSS volume through either of the following methods: · Use a volume directly · Use a volume through a PV or PVC	Not supported

3.4.3.2 Use Apsara Stack disks

You can use Apsara Stack disks to create volumes.

You can use volumes created from Apsara Stack disks in Kubernetes clusters.

Currently, Apsara Stack disks can be mounted to Kubernetes clusters as follows:

- *Static volumes*

  You can use static volumes in either of the following ways:

  - *Directly as volumes*
  - *Through PVs and PVCs*
- *Dynamic volumes*

Notes

- **A disk is a non-shared storage device that can only be mounted to one Pod at one time.**
- **To use a disk as a volume, you must have created the disk and obtained its disk ID.**

  Your disk must meet the following capacity requirements:

  - **A basic disk must have a minimum capacity of 5 GiB.**
  - **An ultra disk must have a minimum capacity of 20 GiB.**
  - **An SSD disk must have a minimum capacity of 20 GiB.**
- **volumeId: The ID of the mounted disk, which must be the same as volumeName and PV Name.**
- **Disks can only be mounted to nodes that are located in the same zone.**
- **Only pay-as-you-go disks can be mounted. If you change the billing method of an ECS instance in the cluster from pay-as-you-go to subscription, you cannot change the billing method of its disks to subscription. Otherwise, the disks cannot be mounted to the cluster.**

Static volumes

You can use Apsara Stack disks as volumes or by creating PVs and PVCs.

Prerequisites

You must have created a disk in the ECS console.

- **Use a disk as a volume**

  Use the following *disk-deploy.yaml* file to create a Pod.

```
apiVersion: extensions/v1beta1
kind: Deployment
metadata:
  name: nginx-disk-deploy
spec:
  replicas: 1
  template:
    metadata:
      labels:
        app: nginx
    spec:
      containers:
      - name: nginx-flexvolume-disk
        image: nginx
        volumeMounts:
          - name: "d-bp1j17ifxfasvts3tf40"
            mountPath: "/data"
      volumes:
        - name: "d-bp1j17ifxfasvts3tf40"
          flexVolume:
```

```
            driver: "alicloud/disk"
            fsType: "ext4"
            options:
              volumeId: "d-bp1j17ifxfasvts3tf40"
```

- **Use a disk to create a PV and a PVC**

  1. **Create a PV**

     You can create a PV through the console or by using a YAML file.

     - **Create a PV by using a YAML file**

       Use the following `disk-pv.yaml` file to create a PV.

       Note:
       The PV name must be the same as the disk ID.

```
apiVersion: v1
kind: PersistentVolume
metadata:
  name: d-bp1j17ifxfasvts3tf40
  labels:
    failure-domain.beta.kubernetes.io/zone: cn-hangzhou-b
    failure-domain.beta.kubernetes.io/region: cn-hangzhou
spec:
  capacity:
    storage: 20Gi
  storageClassName: disk
  accessModes:
    - ReadWriteOnce
  flexVolume:
    driver: "alicloud/disk"
    fsType: "ext4"
    options:
```

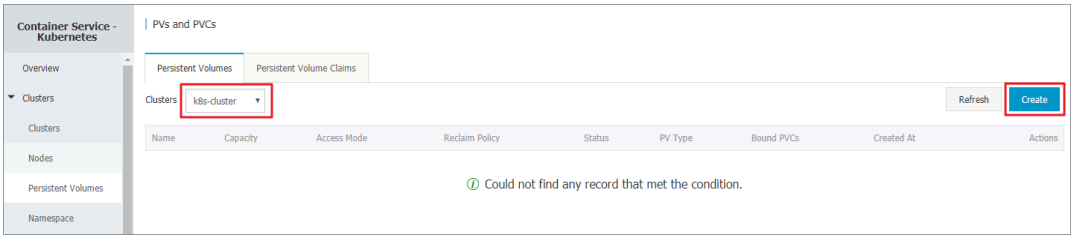
```
        volumeId: "d-bp1j17ifxfasvts3tf40"
```

- **Create a PV through the console**

  a. *Log on to the Container Service console*.

  b. **In the left-side navigation pane, choose Clusters > Persistent Volumes to go to the PVs and PVCs page.**

  c. **On the Persistent Volumes tab, select a cluster and click Create in the upper-right corner.**

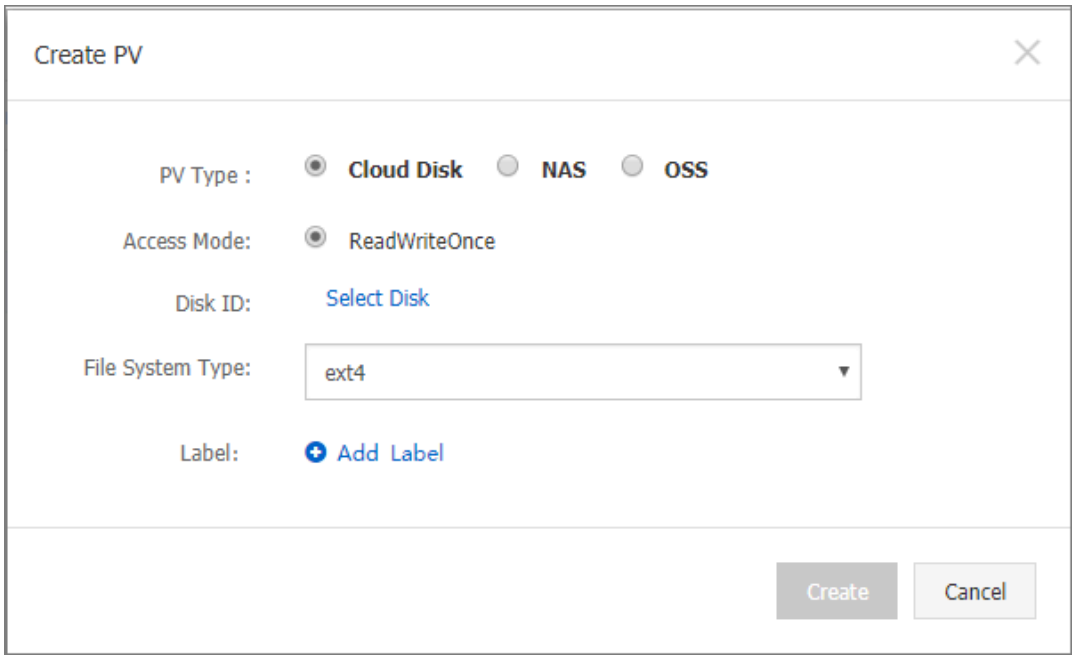


  d. **Set the parameters in the Create PV dialog box.**

Table 3-3: PV configuration

Parameter	Description
PV Type	**In this example, set the PV type to Cloud Disk.**
Access Mode	**Default is ReadWriteOnce.**
Disk ID	**Select a disk that is in the same region and zone as your cluster.**
File System Type	**You can select the data type in which data is stored in the disk. Supported data types include ext4 , ext3, xfs, and vfat. The default setting is ext4.**

Parameter	Description
Label	Add a label for the PV.



e. After the configuration is complete, click Create.

2. Create a PVC

   Use the following `disk-pvc.yaml` file to create a PVC.

```
kind: PersistentVolumeClaim
apiVersion: v1
metadata:
  name: pvc-disk
spec:
  accessModes:
    - ReadWriteOnce
  storageClassName: disk
  resources:
    requests:
      storage: 20Gi
```

3. Create a Pod

   Use the following *disk-pod.yaml* file to create a Pod.

```
apiVersion: v1
kind: Pod
metadata:
  name: "flexvolume-alicloud-example"
spec:
  containers:
    - name: "nginx"
      image: "nginx"
      volumeMounts:
        - name: pvc-disk
          mountPath: "/data"
```

```
  volumes:
  - name: pvc-disk
    persistentVolumeClaim:
      claimName: pvc-disk
```

Dynamic volumes

To use a dynamic volume, you need to manually create a StorageClass, and use storageClassName to specify the disk type in a PVC.

1. **Create a StorageClass**

```
kind: StorageClass
apiVersion: storage.k8s.io/v1beta1
metadata:
  name: alicloud-disk-common-hangzhou-b
provisioner: alicloud/disk
parameters:
  type: cloud_ssd
  regionid: cn-hangzhou
  zoneid: cn-hangzhou-b
```

Parameters:

- **provisioner: Set this parameter to alicloud/disk, which indicates that the Alibaba Cloud Provsioner plugin is used to create the StorageClass.**
- **type: The disk type, which supports the following values: cloud, cloud_effi ciency, cloud_ssd, and available. If you set this parameter to available, the system will try to create a disk in the following order: ultra disk, SSD disk, and basic disk, and will stop trying until a disk is created.**
- **regionid: The region of the disk.**
- **zoneid: The zone of the disk.**

2. **Create a Service**

```
kind: PersistentVolumeClaim
apiVersion: v1
metadata:
  name: disk-common
spec:
  accessModes:
    - ReadWriteOnce
  storageClassName: alicloud-disk-common-hangzhou-b
  resources:
    requests:
      storage: 20Gi
---
kind: Pod
apiVersion: v1
metadata:
  name: disk-pod-common
spec:
  containers:
  - name: disk-pod
```

```
    image: nginx
    volumeMounts:
      - name: disk-pvc
        mountPath: "/mnt"
  restartPolicy: "Never"
  volumes:
    - name: disk-pvc
      persistentVolumeClaim:
        claimName: disk-common
```

Default options

By default, Kubernetes clusters provide the following types of StorageClasses:

- **alicloud-disk-common: basic disk**
- **alicloud-disk-efficiency: ultra disk**
- **alicloud-disk-ssd: SSD disk**
- **alicloud-disk-available: This option ensures high availability. The system will try to create an ultra disk first. If no ultra disk is available in the specified zone, the system will try to create an SSD disk. If no SSD disk is available, the system will try to create a basic disk.**

3. **Create a multi-instance StatefulSet by using a disk**

   We recommend that you use volumeClaimTemplates to create multi-instance StatefulSets. This will dynamically create multiple PVCs and PVs, and bind them together.

```
apiVersion: v1
kind: Service
metadata:
  name: nginx
  labels:
    app: nginx
spec:
  ports:
  - port: 80
    name: web
  clusterIP: None
  selector:
    app: nginx
---
apiVersion: apps/v1beta2
kind: StatefulSet
metadata:
  name: web
spec:
  selector:
    matchLabels:
      app: nginx
  serviceName: "nginx"
  replicas: 2
  template:
    metadata:
      labels:
```

```
        app: nginx
    spec:
      containers:
      - name: nginx
        image: nginx
        ports:
        - containerPort: 80
          name: web
        volumeMounts:
        - name: disk-common
          mountPath: /data
  volumeClaimTemplates:
  - metadata:
      name: disk-common
    spec:
      accessModes: [ "ReadWriteOnce" ]
      storageClassName: "alicloud-disk-common"
      resources:
        requests:
          storage: 10Gi
```

3.4.3.3 Use Apsara Stack NAS volumes

You can use Apsara Stack NAS volumes in Kubernetes clusters.

Currently, Apsara Stack NAS can be mounted to Kubernetes clusters as follows:

- *Static volumes*

  You can use static volumes in either of the following ways:

  - **Through the flexvolume plug-in**
    - **Directly as volumes**
    - **Through PVs and PVCs**
  - **Through the NFS driver**

- *Dynamic volumes*

Prerequisites

You have created a file system in the NAS console and added a mount point. The mount point is used to mount the file system to the Kubernetes cluster. The NAS file system and your cluster are deployed in the same VPC network.

Static volumes

You can use Apsara Stack NAS through the flexvolume plug-in provided by Alibaba Cloud or the NFS driver provided by Kubernetes.

Through the flexvolume plug-in

With the flexvolume plug-in, you can use Apsara Stack NAS volumes directly or by creating PVs and PVCs.

Note:

- **NAS is a shared storage system that can provide storage services to multiple Pods at the same time.**
- **server: The mount point of the NAS volume.**
- **path: The mount directory that connects to the NAS volume. Sub-directories are supported. If no sub-directory exists, the system automatically creates a sub-directory and mounts the NAS volume to the sub-directory.**
- **vers: The version of the NFS mounting protocol. Version 4.0 is supported.**
- **mode: The access permission on the mount directory. Note that the access permission cannot be set if the mount directory is the root directory of the NAS file system. If you set the mode parameter for a NAS file system that stores a large amount of data, the process of mounting the NAS file system to a cluster may take an excessive amount of time or even fail.**

Use a static volume directly

Use the following `nas-deploy.yaml` file to create a Pod.

```
apiVersion: v1
kind: Pod
metadata:
  name: "flexvolume-nas-example"
spec:
  containers:
    - name: "nginx"
      image: "nginx"
      volumeMounts:
        - name: "nas1"
          mountPath: "/data"
  volumes:
    - name: "nas1"
      flexVolume:
        driver: "alicloud/nas"
        options:
          server: "0cd8b4a576-grs79.cn-hangzhou.nas.aliyuncs.com"
          path: "/k8s"
          vers: "4.0"
```

Use a static volume to create a PV and a PVC

Step 1: Create a PV

You can create a PV through the console or by using a YAML file.

- **Create a PV by using a YAML file**

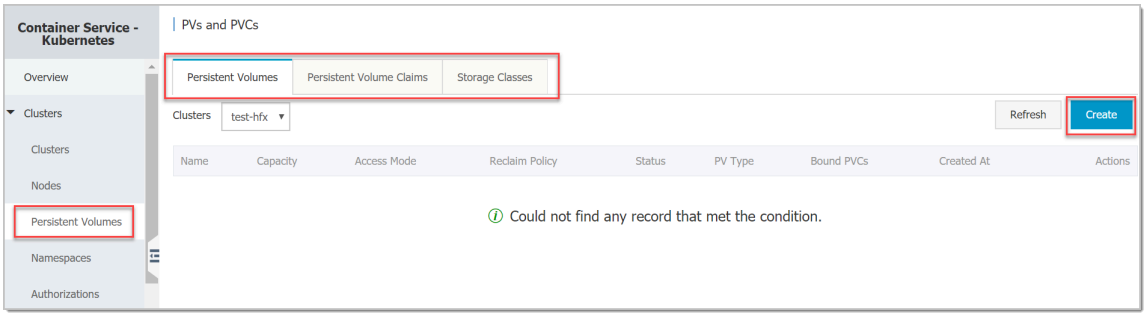
  Use the following `nas-pv.yaml` file to create a PV.

```
apiVersion: v1
kind: PersistentVolume
metadata:
  name: pv-nas
spec:
  capacity:
    storage: 5Gi
  storageClassName: nas
  accessModes:
    - ReadWriteMany
  flexVolume:
    driver: "alicloud/nas"
    options:
      server: "0cd8b4a576-uih75.cn-hangzhou.nas.aliyuncs.com"
      path: "/k8s"
```

```
        vers: "4.0"
```

- Create a PV through the console

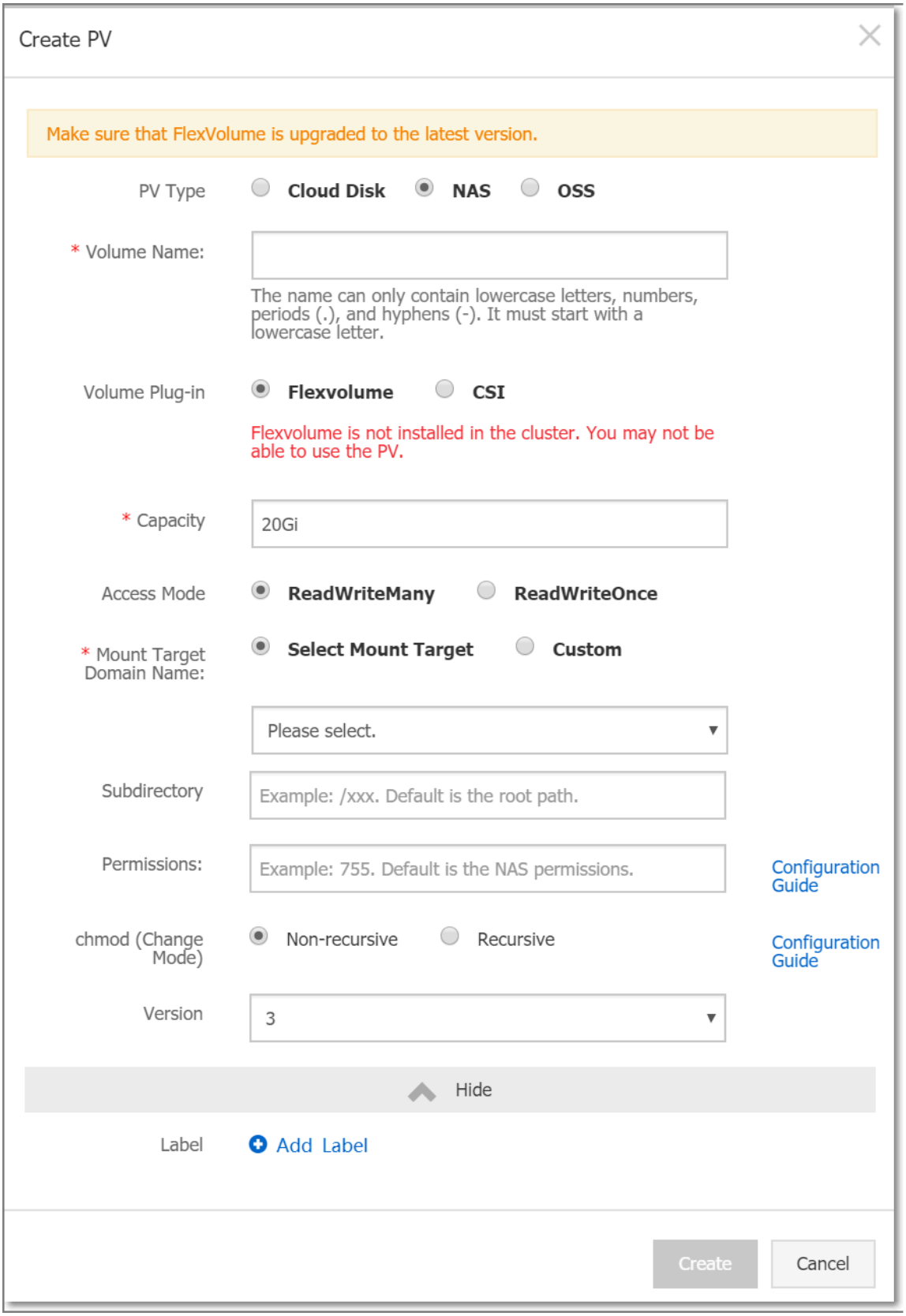
1. Log on to the Container Service console. For more information about logging on to the console, see the Log on to the console section in the Apsara Stack console User Guide.
2. In the left-side navigation pane, choose Clusters > Persistent Volumes to go to the PVs and PVCs page.
3. On the Persistent Volumes tab, select a cluster and click Create in the upper-right corner.



4. In the dialog box that appears, set the following parameters.
   - PV Type: In this example, select NAS.
   - Volume Name: The name of the PV. The name must be unique in the cluster. In this example, enter pv-nas.
   - Capacity: The capacity of the PV. Note that the capacity cannot exceed the capacity of the NAS file system.
   - Access Mode: Default is ReadWriteMany.
   - Mount Point Domain Name: Enter the address of the mount point that is used to mount the NAS file system to the cluster.
   - Subpath: The sub-directory under the NAS file system, which starts with a forward slash (/). When specified, the PV is mounted to the sub-directory.
     - If the sub-directory does not exist under the root directory of the NAS file system, the system automatically creates the sub-directory.
     - This parameter is optional. The PV is mounted to the root directory of the NAS file system by default.
   - Permissions: Set the access permission on the mount directory. For example, 755, 644, or 777.

- **You can set this parameter only if you mount the PV to a sub-directory of the NAS file system. You cannot set permissions on the root directory of the NAS file system.**
- **This parameter is optional. The original permissions are used by default.**

- **Label: Add labels for the PV.**



5. **After the configuration is complete, click Create.**

Step 2: Create a PVC

Use the following *nas-pvc.yaml* file to create a PVC.

```
kind: PersistentVolumeClaim
apiVersion: v1
metadata:
  name: pvc-nas
spec:
  accessModes:
    - ReadWriteMany
  storageClassName: nas
  resources:
    requests:
      storage: 5Gi
```

Step 3: Create a Pod

Use the following *nas-pod.yaml* file to create a Pod.

```
apiVersion: v1
kind: Pod
metadata:
  name: "flexvolume-nas-example"
spec:
  containers:
    - name: "nginx"
      image: "nginx"
      volumeMounts:
          - name: pvc-nas
            mountPath: "/data"
  volumes:
  - name: pvc-nas
    persistentVolumeClaim:
        claimName: pvc-nas
```

Through the NFS driver

Step 1: Create a NAS file system

Log on to the NAS console. For more information about logging on to the console, see the Create NAS file systems section in the Apsara Stack console User Guide.

Note:

The NAS file system and your cluster must be deployed in the same region.

Assume that the mount point of the NAS file system is `055f84ad83-ixxxx.cn-hangzhou.nas.aliyuncs.com`.

Step 2: Create a PV

You can create a PV by using an orchestration template or through the Container Service console.

- **Use an orchestration template**

  Use the *nas-pv.yaml* file to create a PV.

  Run the following command to create a PV from the NAS file system:

```
root@master # cat << EOF |kubectl apply -f -
apiVersion: v1
kind: PersistentVolume
metadata:
   name: nas
spec:
   capacity:
     storage: 8Gi
   accessModes:
     - ReadWriteMany
   persistentVolumeReclaimPolicy: Retain
   nfs:
     path: /
     server: 055f84ad83-ixxxx.cn-hangzhou.nas.aliyuncs.com
EOF
```

- **Create a PV through the console**

  For more information, see *Use a static volume to create a PV and a PVC*.

Step 3: Create a PVC

Create a PVC and associate it with the PV created from the previous step.

```
root@master # cat << EOF | kubectl apply -f -
apiVersion: v1
kind: PersistentVolumeClaim
metadata:
     name: nasclaim
spec:
     accessModes:
       - ReadWriteMany
     resources:
       requests:
         storage: 8Gi
EOF
```

Step 4: Create a Pod

Create an application to use and mount the PV.

```
root@master # cat << EOF |kubectl apply -f -
apiVersion: v1
kind: Pod
metadata:
     name: mypod
spec:
     containers:
       - name: myfrontend
         image: registry.aliyuncs.com/spacexnice/netdia:latest
         volumeMounts:
         - mountPath: "/var/www/html"
           name: mypd
```

```
      volumes:
        - name: mypd
          persistentVolumeClaim:
            claimName: nasclaim
EOF
```

The NAS file system is now mounted to the application that runs on the Pod.

Dynamic volumes

To use dynamic NAS volumes, you need to manually install a driver plug-in and configure a NAS mount point.

Install the plug-in

```
apiVersion: storage.k8s.io/v1
kind: StorageClass
metadata:
  name: alicloud-nas
provisioner: alicloud/nas
---
apiVersion: v1
kind: ServiceAccount
metadata:
  name: alicloud-nas-controller
  namespace: kube-system
---
kind: ClusterRoleBinding
apiVersion: rbac.authorization.k8s.io/v1beta1
metadata:
  name: run-alicloud-nas-controller
subjects:
  - kind: ServiceAccount
    name: alicloud-nas-controller
    namespace: kube-system
roleRef:
  kind: ClusterRole
  name: alicloud-disk-controller-runner
  apiGroup: rbac.authorization.k8s.io
---
kind: Deployment
apiVersion: extensions/v1beta1
metadata:
  name: alicloud-nas-controller
  namespace: kube-system
spec:
  replicas: 1
  strategy:
    type: Recreate
  template:
    metadata:
      labels:
        app: alicloud-nas-controller
    spec:
      tolerations:
      - effect: NoSchedule
        operator: Exists
        key: node-role.kubernetes.io/master
      - effect: NoSchedule
        operator: Exists
        key: node.cloudprovider.kubernetes.io/uninitialized
```

```
      nodeSelector:
         node-role.kubernetes.io/master: ""
      serviceAccount: alicloud-nas-controller
      containers:
        - name: alicloud-nas-controller
          image: registry.cn-hangzhou.aliyuncs.com/acs/alicloud-nas-
controller:v1.8.4
          volumeMounts:
          - mountPath: /persistentvolumes
            name: nfs-client-root
          env:
            - name: PROVISIONER_NAME
              value: alicloud/nas
            - name: NFS_SERVER
              value: 0cd8b4a576-mmi32.cn-hangzhou.nas.aliyuncs.com
            - name: NFS_PATH
              value: /
      volumes:
        - name: nfs-client-root
          nfs:
            server: 0cd8b4a576-mmi32.cn-hangzhou.nas.aliyuncs.com
            path: /
```

Use dynamic volumes

```
apiVersion: apps/v1beta1
kind: StatefulSet
metadata:
  name: web
spec:
  serviceName: "nginx"
  replicas: 2
  volumeClaimTemplates:
  - metadata:
      name: html
    spec:
      accessModes:
        - ReadWriteOnce
      storageClassName: alicloud-nas
      resources:
        requests:
          storage: 2Gi
  template:
    metadata:
      labels:
        app: nginx
    spec:
      containers:
      - name: nginx
        image: nginx:alpine
        volumeMounts:
        - mountPath: "/usr/share/nginx/html/"
          name: html
```

3.4.3.4 Use Apsara Stack OSS volumes

You can use Apsara Stack OSS volumes to create PVs in Kubernetes clusters.

You can use OSS volumes in either of the following ways:

- **Directly as volumes**

- **Through PVs or PVCs**

Prerequisites

You have created a bucket in the OSS console.

Background

- **Currently, you can only use static OSS volumes.**
- **OSS is a shared storage system that can provide storage services to multiple Pods at the same time.**
- **bucket: Only buckets can be mounted to a Kubernetes cluster. The sub-directories or files under a bucket cannot be mounted to a Kubernetes cluster.**
- **url: The OSS endpoint, namely, the domain that is used to mount an OSS bucket to a cluster.**
- **akId: Your AccessKey ID.**
- **akSecret: Your AccessKey Secret.**
- **otherOpts: Enter custom parameters in the format of `-o *** -o ***`.**

Note:

To use OSS volumes, you must create a Secret and enter your AccessKey information in the Secret when you deploy the flexvolume service.

Use static OSS volumes

- **Use an OSS bucket as a volume**

  Use the following *oss-deploy.yaml* file to create a Pod.

```
apiVersion: extensions/v1beta1
kind: Deployment
metadata:
  name: nginx-oss-deploy
spec:
  replicas: 1
  template:
    metadata:
      labels:
        app: nginx
    spec:
      containers:
      - name: nginx-flexvolume-oss
        image: nginx
        volumeMounts:
          - name: "oss1"
            mountPath: "/data"
      volumes:
        - name: "oss1"
          flexVolume:
```

```
            driver: "alicloud/oss"
            options:
              bucket: "docker"
              url: "oss-cn-hangzhou.aliyuncs.com"
              akId: ***
              akSecret: ***
              otherOpts: "-o max_stat_cache_size=0 -o allow_other"
```

- **Use an OSS bucket to create a PV and a PVC**

  1. **Create a PV**

     You can create a PV through the console or by using a YAML file.

     - **Create a PV by using a YAML file**

       Use the following *oss-pv.yaml* **file to create a PV.**

       ```
       apiVersion: v1
       kind: PersistentVolume
       metadata:
         name: pv-oss
       spec:
         capacity:
           storage: 5Gi
         accessModes:
           - ReadWriteMany
         storageClassName: oss
         flexVolume:
           driver: "alicloud/oss"
           options:
             bucket: "docker"
             url: "oss-cn-hangzhou.aliyuncs.com"
             akId: ***
             akSecret: ***
       ```

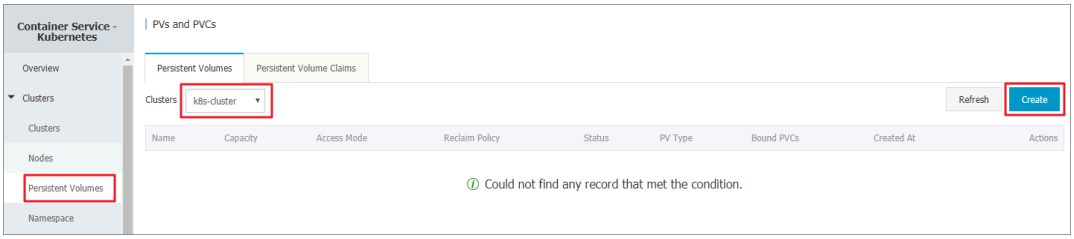
```
      otherOpts: "-o max_stat_cache_size=0 -o allow_other"
```

- **Create a PV through the console**

  a. *Log on to the Container Service console*.

  b. **In the left-side navigation pane, choose Clusters > Persistent Volumes to go to the PVs and PVCs page.**

  c. **On the Persistent Volumes tab, select a cluster and click Create in the upper-right corner.**

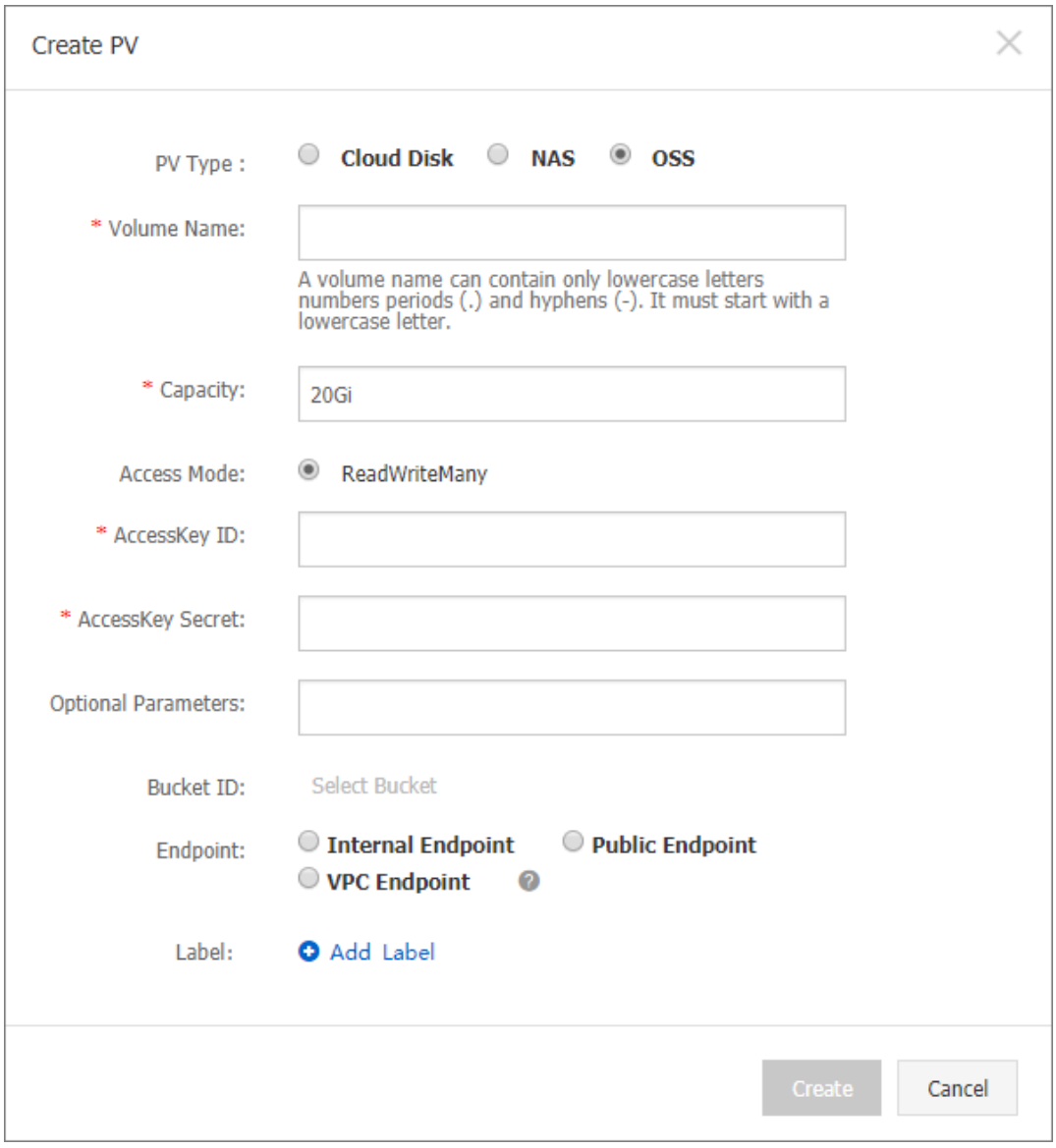


  d. **Set the parameters in the Create PV dialog box.**

Table 3-4: Create a PV through the console

Parameter	Description
PV Type	In this example, select OSS.
Volume Name	The name of the PV. The name must be unique in the cluster. In this example, enter pv-oss.
Capacity	The capacity of the PV.
Access Mode	Default is ReadWriteMany.
AccessKey ID and AccessKey Secret	The AccessKey that is required to access the OSS. You can choose User Center > Department Management in the Apsara Stack console, select the current department, and click Obtain Department AccessKey to obtain the AccessKey pair.
Bucket ID	The name of the OSS bucket that you want to use. Click Select Bucket. In the dialog box that appears, choose the target bucket and click Select.
Endpoint	We recommend that you choose Internal Endpoint.

Parameter	Description
Label	Add labels for the PV.



e. **After the configuration is complete, click Create.**

2. **Create a PVC**

   Use the following *oss-pvc.yaml* **file to create a PVC.**

```
kind: PersistentVolumeClaim
apiVersion: v1
metadata:
  name: pvc-oss
spec:
  storageClassName: oss
  accessModes:
    - ReadWriteMany
  resources:
    requests:
```

```
      storage: 5Gi
```

3. **Create a Pod**

   Use the following *oss-pod.yaml* **file to create a Pod.**

```
apiVersion: v1
kind: Pod
metadata:
  name: "flexvolume-oss-example"
spec:
  containers:
    - name: "nginx"
      image: "nginx"
      volumeMounts:
          - name: pvc-oss
            mountPath: "/data"
  volumes:
  - name: pvc-oss
    persistentVolumeClaim:
        claimName: pvc-oss
```

Use dynamic OSS volumes

Currently, dynamic OSS volumes are not supported.

3.4.3.5 Create PVCs

Prerequisites

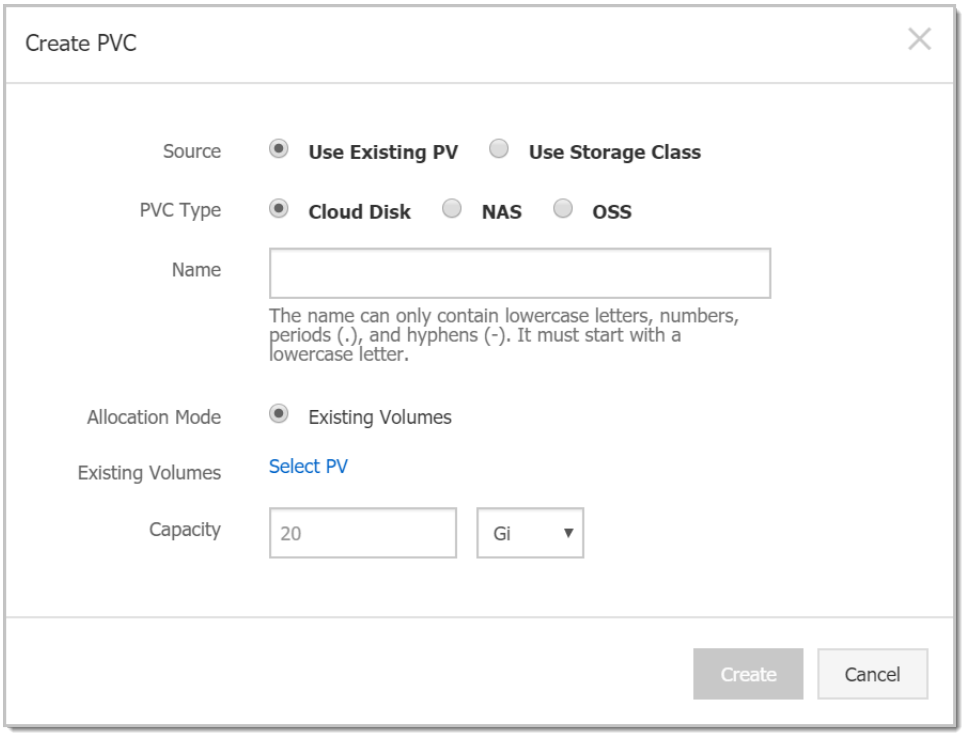
- **You have created a Kubernetes cluster. For more information, see** *Create a Kubernetes cluster*.
- **You have created a PV. This example uses a PV created from a disk. For more information, see** *Use Apsara Stack disks*.

  By default, PVCs are associated with PVs that have the `alicloud-pvname` **label. PVs created through the Container Service console all have this label. If the PV does not have this label, you need to add the label before you can associate it with a PVC.**

Procedure

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Applications > Persistent Volume Claims to go to the PVs and PVCs page.**
3. **On the Persistent Volume Claims tab, select the cluster and namespace, and click Create in the upper-right corner.**

4. Set the parameters in the Create PVC dialog box, and click Create.

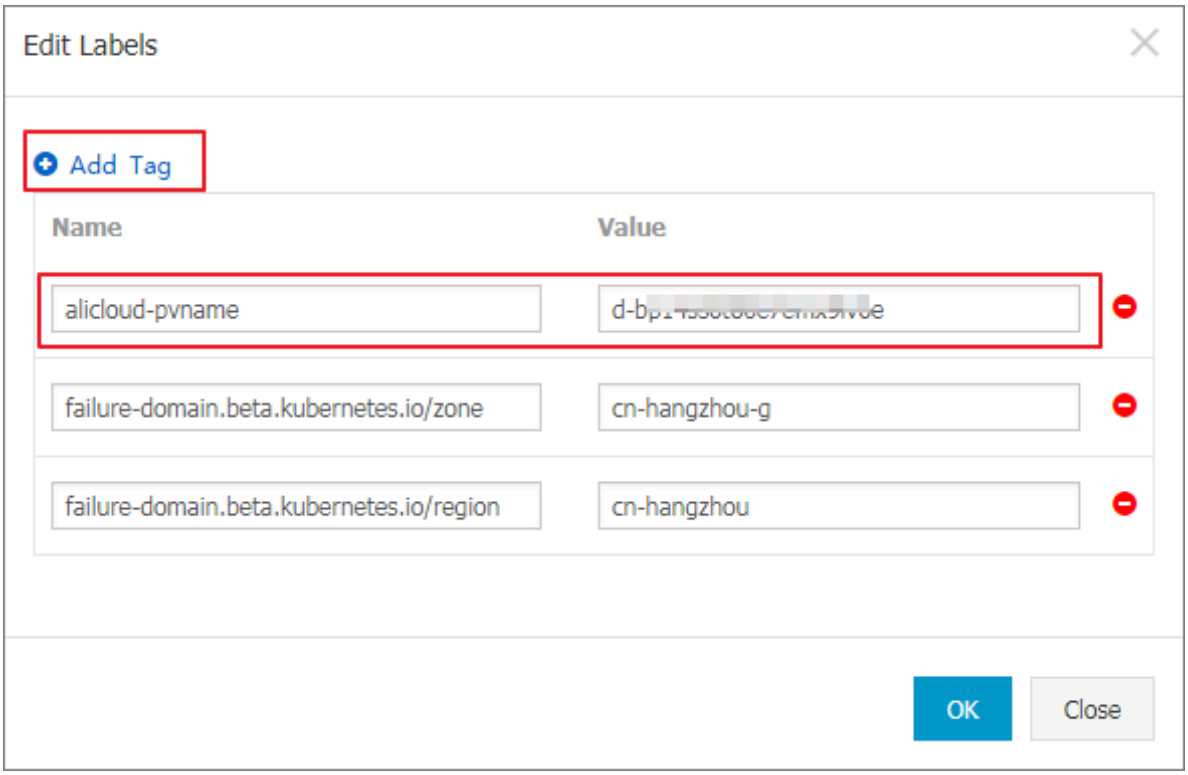


- PVC Type: The same as the types of PVs. Three types, cloud disk, NAS, and OSS, are supported.
- Name: The PVC name.
- Allocation Mode: Currently, only existing volumes are supported.
- Existing Volumes: Select to associate with PVs of the same type.
- Capacity: The claimed usage. Must not be larger than the total capacity of the associated PVs.

Note:
If your cluster already has a PV that is not used, but you cannot find it in the Select PV dialog box, the reason may be that the PV does not have the `alicloud-pvname` label.

If you cannot find available PVs, you can choose Clusters > Persistent Volumes in the left-side navigation pane. Find the PV that you want to use and click Manage Labels in the Actions column. You can add a label for the PV and set the label

name to `alicloud-pvname` and the value to the PV name. By default, the disk ID is used as the PV name if the PV is created from a disk.



5. Go to the Persistent Volume Claims page. You can find the newly created PVC in the list.

3.4.3.6 Use PVCs

You can use persistent volume claims (PVCs) in your applications.

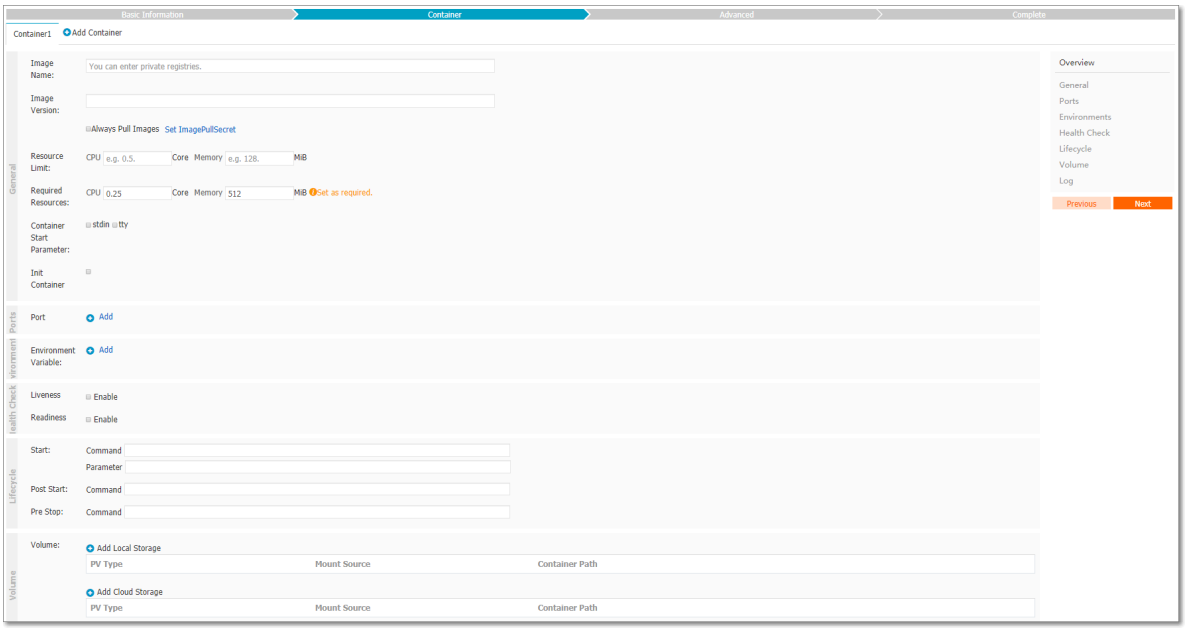
Prerequisites

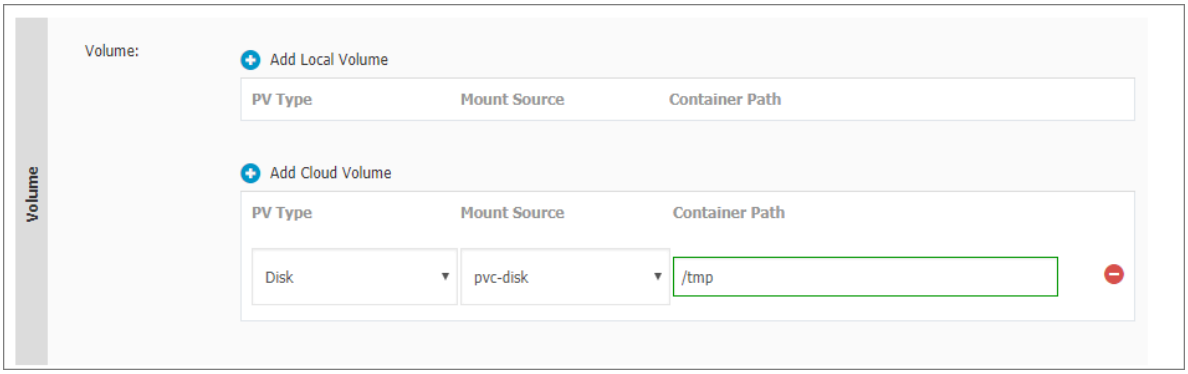
- You have created a Kubernetes cluster. For more information, see *Create a Kubernetes cluster*.
- You have created a PVC. This example uses a PVC named pvc-disk. For more information, see *Create PVCs*.

Procedure

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Applications > Deployments to go to the Deployments page. In the upper-right corner, click Create from Image.
3. On the Basic Information page, specify the application name, cluster, namespace, number of replicas, type, labels, and annotations. Then click Next.

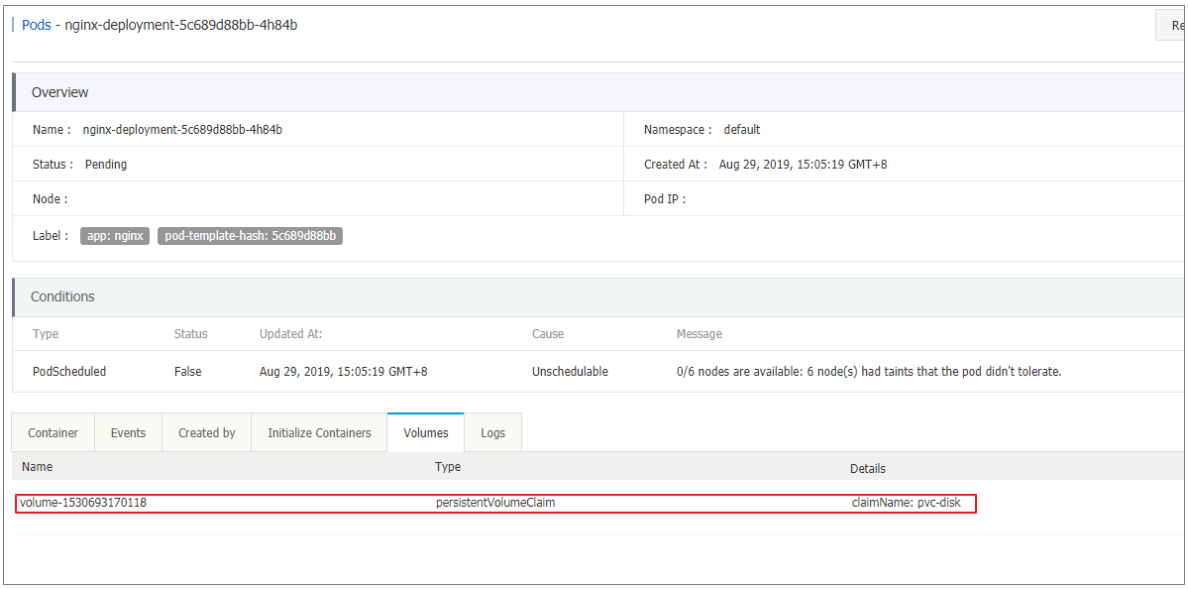
4. On the Container page, select the image. Then, specify the type of the cloud volume. Currently, cloud disk, NAS, and OSS are supported. In this example, select the pvc-disk PVC and click Next.





5. Configure the test-nginx application, and click Create.
6. After the application is created, click Applications > Pods in the left-side navigation pane. Select the Pod where the application belongs to, and click View Details.

7. On the Pod details page, click Volumes and the Pod is associated with the pvc-disk PVC.



3.4.4 Namespaces

3.4.4.1 Create a namespace

This topic describes how to create a namespace through the Container Service console.

Prerequisites

You have created a Kubernetes cluster.

Context

You can use namespaces to create multiple virtual spaces in a Kubernetes cluster. When a large number of users share a cluster, multiple namespaces can be used to effectively divide different workspaces and assign cluster resources to different tasks. Furthermore, you can use *resource-quotas* to assign resources to each namespace.

Procedure

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Clusters > Namespaces. The Namespaces page appears.
3. Select a cluster from the Cluster drop-down list, and then click Create in the upper-right corner.

4. **In the dialog box that appears, set the namespace parameters.**

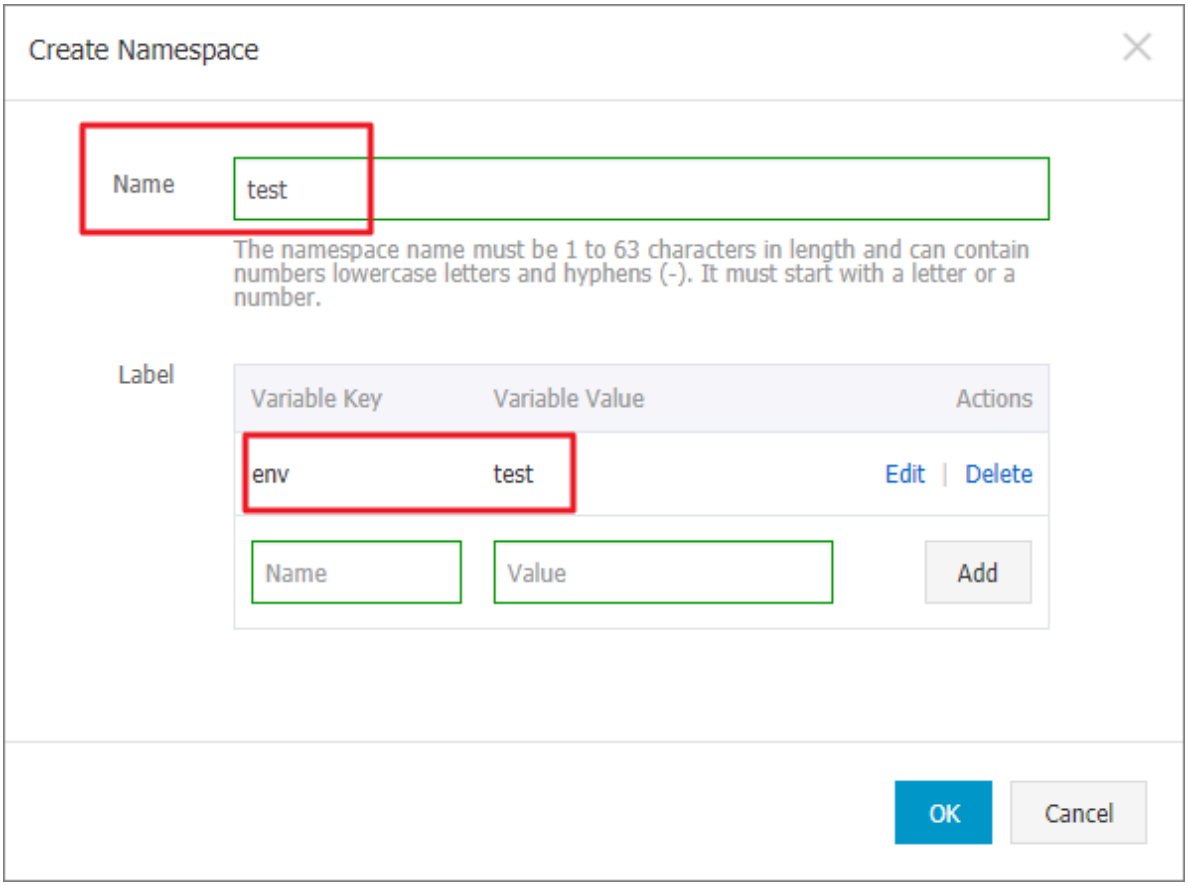
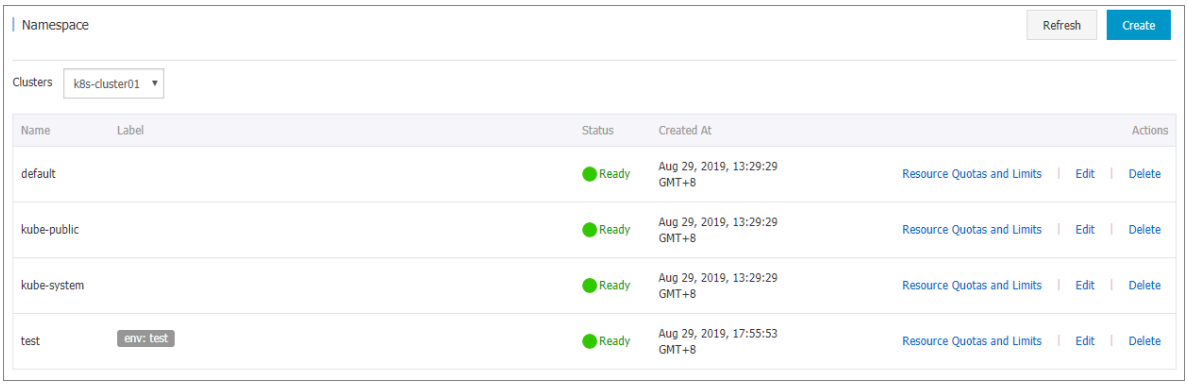


Table 3-5: Create a namespace

Parameter	Description
Name	Enter a name for the namespace. The name must be 1 to 63 characters in length and can contain digits, letters, and hyphens (-). It must start and end with a letter or digit. In this example, test is used as the name.

Parameter	Description
Labels	Labels: Add one or multiple labels to the namespace to indicate its characteristics. For example, you can set a label to indicate that this namespace is used for the test environment. You can enter a variable name and a variable value, and then click Add on the right to add a label to the namespace.

5. After you complete the settings, click OK.
6. The namespace named test is displayed on the Namespaces page.



3.4.4.2 Set resource quotas and limits for a namespace

This topic describes how to set resource quotas and limits for a namespace through the Container Service console.

Prerequisites

- You have created a Kubernetes cluster. For more information, see *Create a Kubernetes cluster*.
- You have created a namespace. In this topic, a namespace named test is used. For more information, see *Create a namespace*.
- You have connected to the master node of the cluster. For more information, see *Connect to a Kubernetes cluster through kubectl*.

Context

By default, there are no limits to the amount of node CPU and memory resources that can be occupied by a single running pod. Because of this, a single pod within a namespace may deplete the resources of the entire cluster.

Namespaces can be used as virtual clusters to serve multiple users. Therefore, setting resource quotas for a namespace is regarded as a best practice.

You can set the quotas for namespace resources, such as the CPU, memory, and number of pods. For more information, see *Resource quotas*.

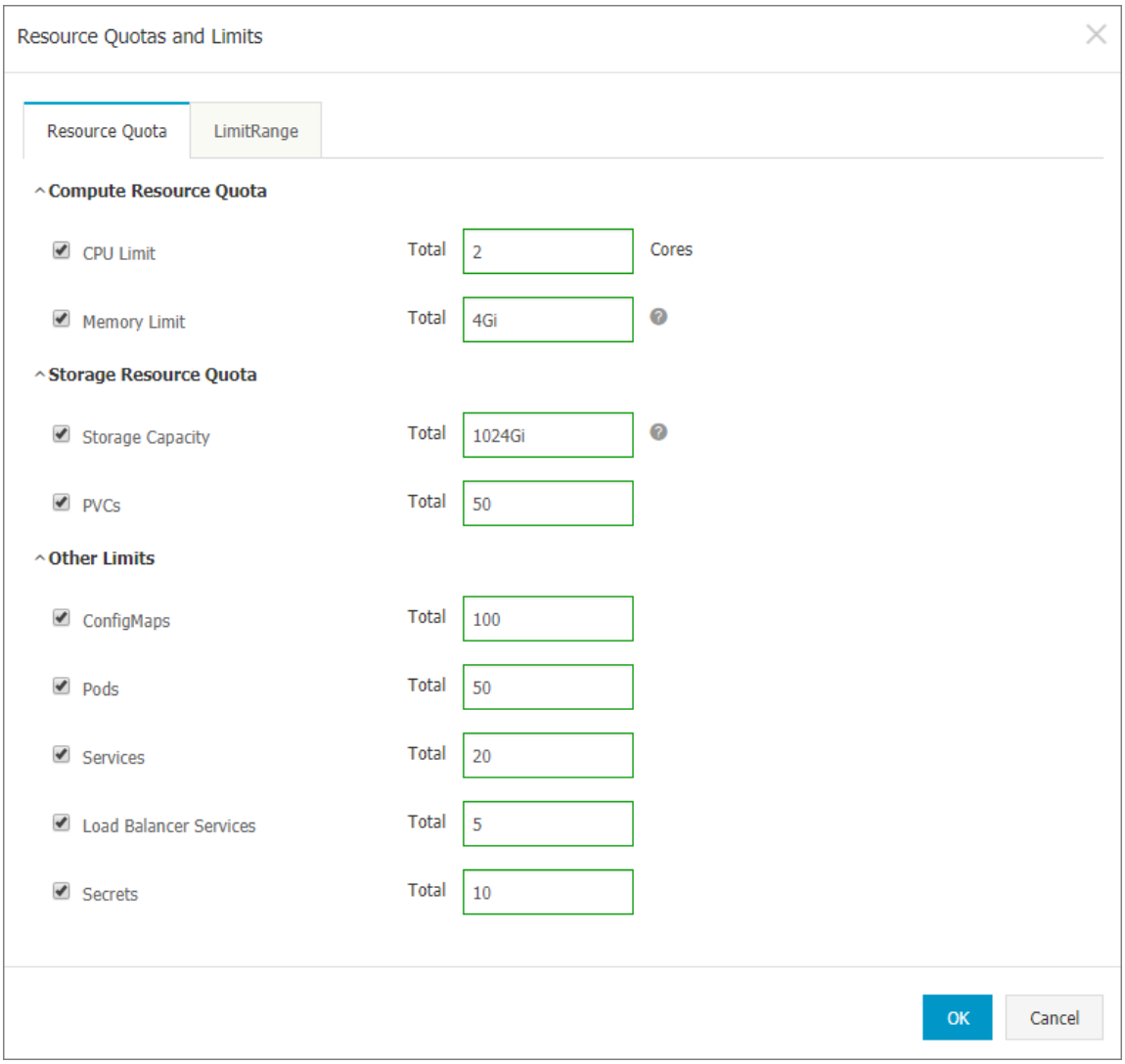
Procedure

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Clusters > Namespaces. Select a cluster from the Cluster drop-down list. Click ResourceQuota and LimitRange corresponding to a namespace.
3. In the ResourceQuota and LimitRange dialog box, set the resource quotas and default resource limits.

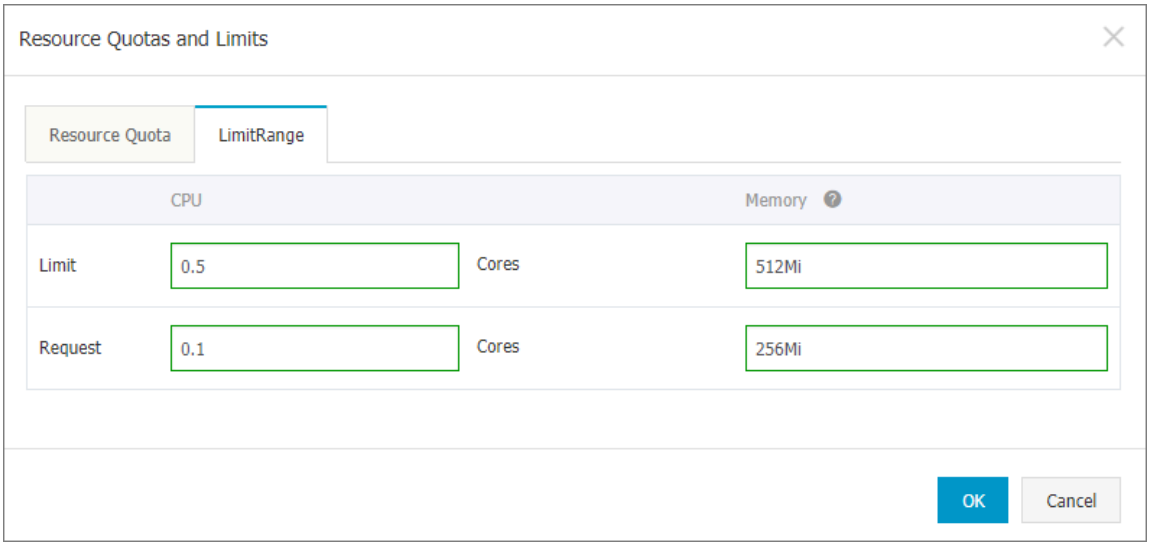
Note:

After setting CPU and memory quotas for a namespace, you must specify CPU and memory resource limits or set the default resource limits for the namespace when creating a pod. For more information, see *Resource quotas*.

a) Set resource quotas for the namespace.



b) To control the amount of resources consumed by containers, set resource limits and resource requests for containers in this namespace. For more information, see *https://kubernetes.io//memory-default-namespace/*.



4. **Connect to the master node and then run the following commands to view the resources of the test namespace:**

```
# kubectl get limitrange,ResourceQuota -n test
NAME AGE
limitrange/limits 8m

NAME AGE
resourcequota/quota 8m


# kubectl describe limitrange/limits resourcequota/quota -n test
Name: limits
Namespace: test
Type Resource Min Max Default Request Default Limit Max Limit/
Request Ratio
---- -------- --- --- --------------- -------------
 -----------------------
Container cpu - - 100m 500m -
Container memory - - 256Mi 512Mi -

Name: quota
Namespace: test
Resource Used Hard
-------- ---- ----
configmaps 0 100
limits.cpu 0 2
limits.memory 0 4Gi
persistentvolumeclaims 0 50
pods 0 50
requests.storage 0 1Ti
secrets 1 10
services 0 20
services.loadbalancers 0 5
```

3.4.4.3 Edit a namespace

This topic describes how to edit a namespace.

Prerequisites

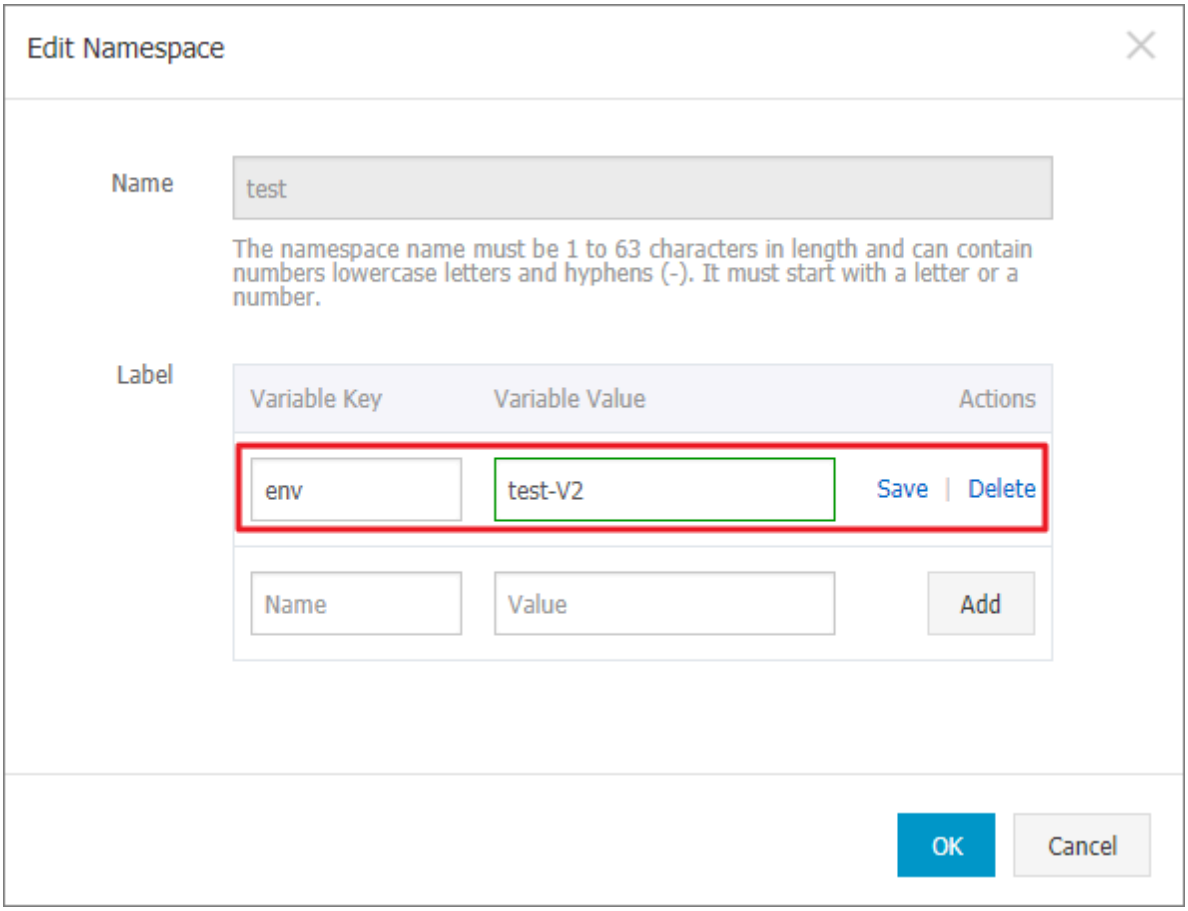
- You have created a Kubernetes cluster. For more information, see *Create a Kubernetes cluster*.
- You have created a namespace. In this topic, a namespace named test is used. For more information, see *Create a namespace*.

Context

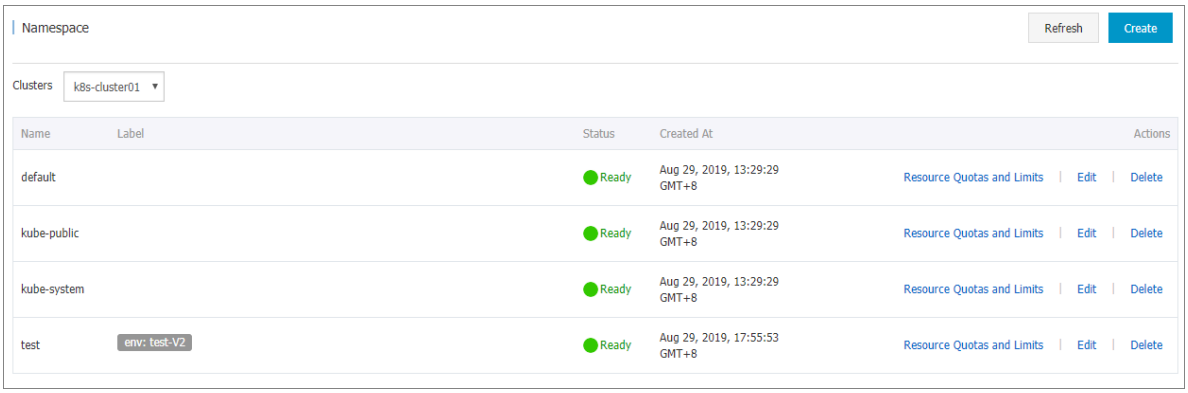
You can add, modify, or delete namespace labels.

Procedure

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Clusters > Namespaces. The Namespaces page appears.
3. Select a cluster from the Cluster drop-down list. Then, click Edit in the Actions column corresponding to a namespace.
4. In the dialog box that appears, click Edit to change the label of the namespace. For this example, change the label to `env:test-V2` and click Save.



5. Click OK. The changed namespace label is then displayed on the Namespaces page.



3.4.4.4 Delete a namespace

This topic describes how to delete a namespace that is no longer required.

Prerequisites

- You have created a Kubernetes cluster. For more information, see *Create a Kubernetes cluster*.
- You have created a namespace. In this topic, a namespace named test is used. For more information, see *Create a namespace*.

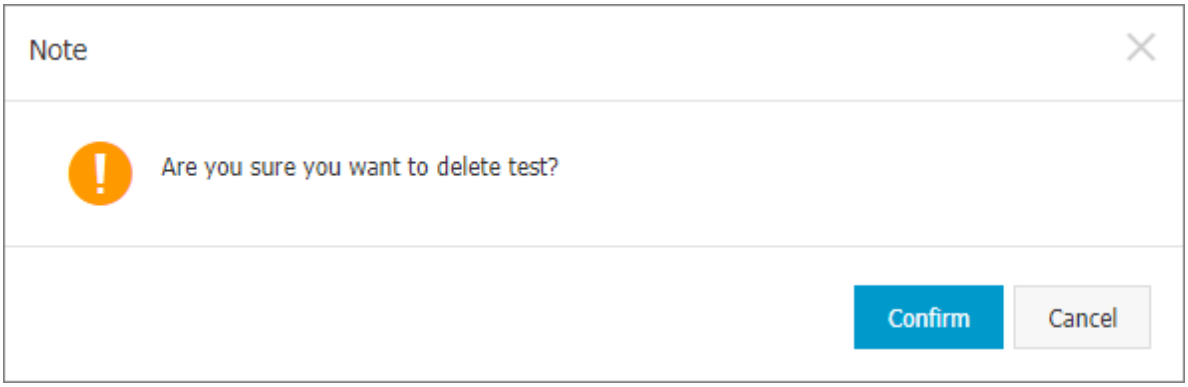
Context

> **Note:**
> Deleting a namespace also deletes all resources objects that belong to the namespace. Exercise caution when performing this action.

Procedure

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Clusters > Namespaces. The Namespaces page appears.
3. Select a cluster from the Cluster drop-down list. Then, click Delete in the Actions column corresponding to a namespace.

4. In the message that appears, click Confirm.



5. The namespace is then deleted from the namespace list, and its resource objects are also deleted.

3.4.5 Applications

3.4.5.1 Create an application from an image

You can use an image to create an Nginx application that is accessible to the Internet.

Prerequisites

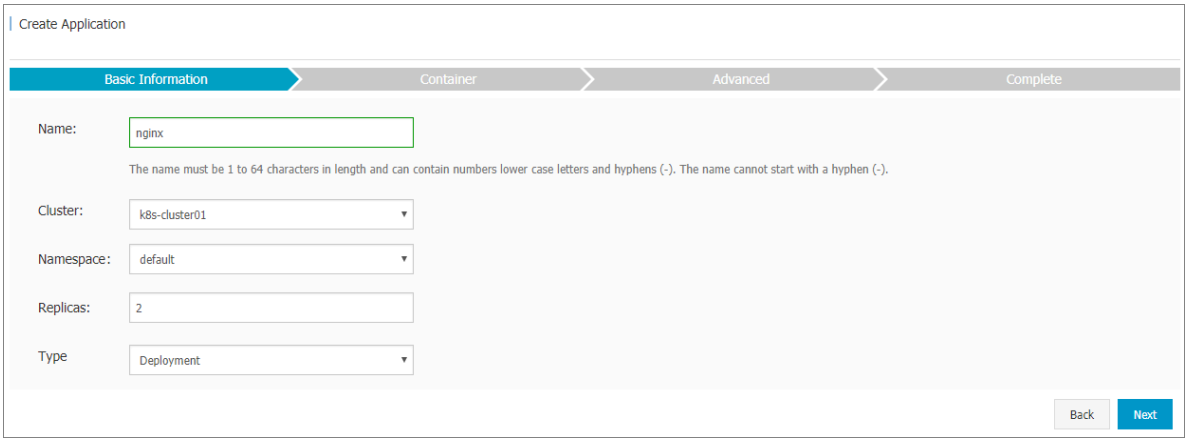
- **You have created a Kubernetes cluster. For more information, see** *Create a Kubernetes cluster*.
- **Your Kubernetes cluster is accessible to the Internet.**

Procedure

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Applications > Deployments and click Create from Image in the upper-right corner.**

3. **Set the following parameters: Name, Cluster, Namespace, Replicas, and Type. Click Next.**

   If you do not set the Namespace parameter, the default namespace is used.



4. **Configure the container.**

   Note:
   You can configure multiple containers for the Pods where the application is deployed.

   a) **Configure the general settings.**

   Table 3-6: Container general settings

Parameter	Description
Image Name	You can click Select Image to select the image in the dialog box that appears. In this example, select Nginx and click OK. You can also enter a private registry to specify the image. The format is as follows: `domainname/namespace/imagename`.
Image Version	You can click Select Image Version to select the version. If you do not specify the image version, the latest version is used by default.

Parameter	Description
Always Pull Image	To improve efficiency, Container Service caches the image. During deployment, if the version of the newly specified image is the same as that of the cached image , Container Service will reuse the cached image rather than pull the image again. Therefore, if the image version is kept unchanged for some reason, for example, to make it easy to run upper-layer services, when the code or image is updated, the previously cached image will be used during deployment. When this option is selected, Container Service will always re-pull the image from the repository to deploy the application. This ensures that the latest image and code are used.
Set Image Secret	Click Set Image Secret to set the image secret. You must set the secret if you need to access a private repository.
Resource Limit	The upper limits of CPU and memory resources that can be used by this application. This prevents the applicatio n from using excessive resources. The unit of CPU resources is Core. The unit of memory is MiB.
Required Resources	The amount of CPU and memory resources that are reserved for this application. These resources are exclusive to the container. This prevents the application from being unavailable when other services or processes compete for resources.
Init Container	When this option is selected, the system creates an Init Container that contains useful tools. For more information, see *https://kubernetes.io/docs/concepts/workloads/pods/init-containers/*.

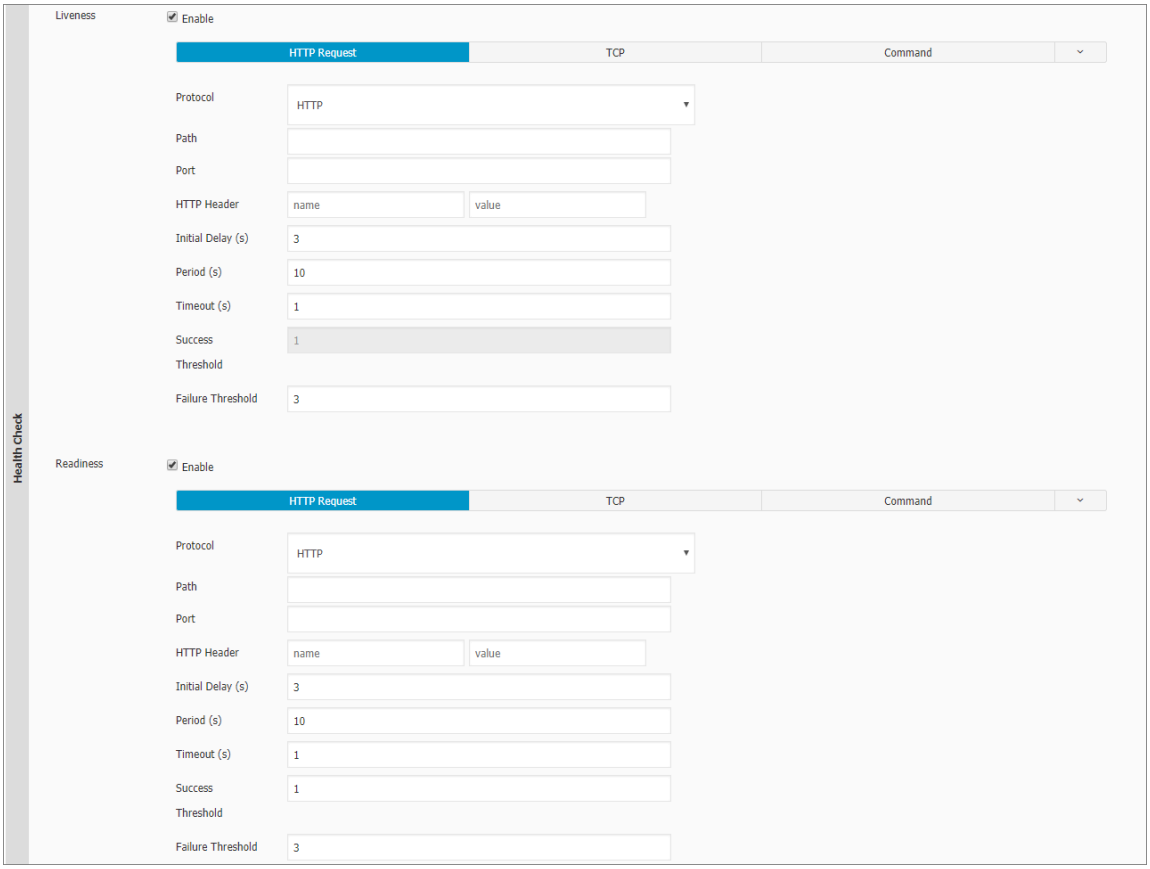
b) Optional: Configure environment variables.

You can configure environment variables for the Pods by using key-value pairs. Environment variables are used to add environment labels or pass configurations to the Pods. For more information, see *Pod variable*.

c) Optional: Configure Health Check settings.

Health check settings include liveness and readiness probes. Liveness probes are used to detect when to restart the container. Readiness probes determine if the container is ready to start accepting traffic. For more information about

health check, see *https://kubernetes.io/docs/tasks/configure-pod-container/configure-liveness-readiness-probes*.



Request type	Description
HTTP request	Sends an HTTP GET request to the container. Supported parameters are as follows: · Protocol: HTTP or HTTPS · Path: The requested path on the server. · Port: The port exposed by the container. The port number must be in the range of 1 to 65535. · HTTP Header: The custom headers in the HTTP request. Replicate headers are allowed. Supports key-value pairs. · Initial Delay (s): The initialDelaySeconds field. The time (in seconds) to wait before performing the first probe after the container is started. Default is 3. · Period (s): The periodSeconds field. How often (in seconds) to perform the probe. Default is 10. Minimum is 1. · Timeout (s): The timeoutSeconds. The time (in seconds) after which the probe times out. Default is 1. Minimum is 1. · Healthy Threshold: The minimum number of consecutive successes that must occur for the probe to be considered successful after having failed. Default is 1. Minimum is 1. For liveness probes, this parameter must be set to 1. · Unhealthy Threshold: The minimum number of consecutive failures that must occur for the probe to be considered failed after having succeeded. Default is 3. Minimum is 1.

Request type	Description
TCP connection	Sends a TCP socket to the container. The kubelet will attempt to open a socket to your container on the specified port. If a connection can be established, the container is considered healthy. Otherwise, it is considered unhealthy. Supported parameters are as follows: · Port: The port exposed by the container. The port number must be in the range of 1 to 65535. · Initial Delay (s): The initialDelaySeconds field. The time (in seconds) to wait before performing the first probe after the container is started. Default is 15. · Period (s): The periodSeconds field. How often (in seconds) to perform the probe. Default is 10. Minimum is 1. · Timeout (s): The timeoutSeconds. The time (in seconds) after which the probe times out. Default is 1. Minimum is 1. · Healthy Threshold: The minimum number of consecutive successes that must occur for the probe to be considered successful after having failed. Default is 1. Minimum is 1. For liveness probes, this parameter must be set to 1. · Unhealthy Threshold: The minimum number of consecutive failures that must occur for the probe to be considered failed after having succeeded. Default is 3. Minimum is 1.

Request type	Description
Command line	Runs a probe command in the container to check its health. Supported parameters are as follows: · Command: The probe command that is used to check the health of the container. · Initial Delay (s): The initialDelaySeconds field. The time (in seconds) to wait before performing the first probe after the container is started. Default is 5. · Period (s): The periodSeconds field. How often (in seconds) to perform the probe. Default is 10. Minimum is 1. · Timeout (s): The timeoutSeconds. The time (in seconds) after which the probe times out. Default is 1. Minimum is 1. · Healthy Threshold: The minimum number of consecutive successes that must occur for the probe to be considered successful after having failed. Default is 1. Minimum is 1. For liveness probes, this parameter must be set to 1. · Unhealthy Threshold: The minimum number of consecutive failures that must occur for the probe to be considered failed after having succeeded. Default is 3. Minimum is 1.

d) Configure lifecycle events.

You can set the following parameters to configure the lifecycle of the container: start, postStart, and preStop. For more information, see *https://kubernetes.io/docs/tasks/configure-pod-container/attach-handler-lifecycle-event/*.

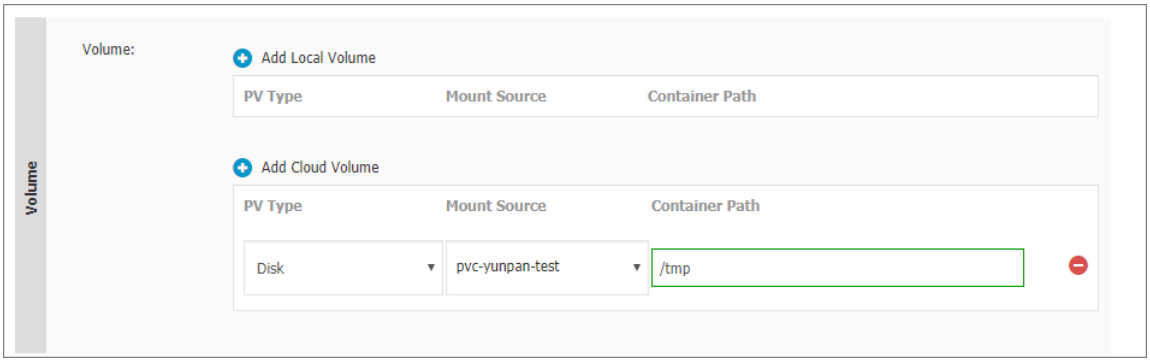
- Start: The pre-start command and parameter.
- Post Start: The post-start command.
- Pre Stop: The pre-stop command.

e) Optional: Configure volumes.

Local volumes and cloud volumes are supported.

- Local Volume: Supports hostPath, ConfigMaps, Secrets, and temporary directories. Local volumes mount the corresponding mount source to a path in the container. For more information, see *Volumes*.
- Cloud Volume: Supports three types of PVs: cloud disks, NAS, and OSS.

This example selects a PV created from a cloud disk and mounts the PV to the `/tmp` path in the container. Data generated in this path is stored to the cloud disk.



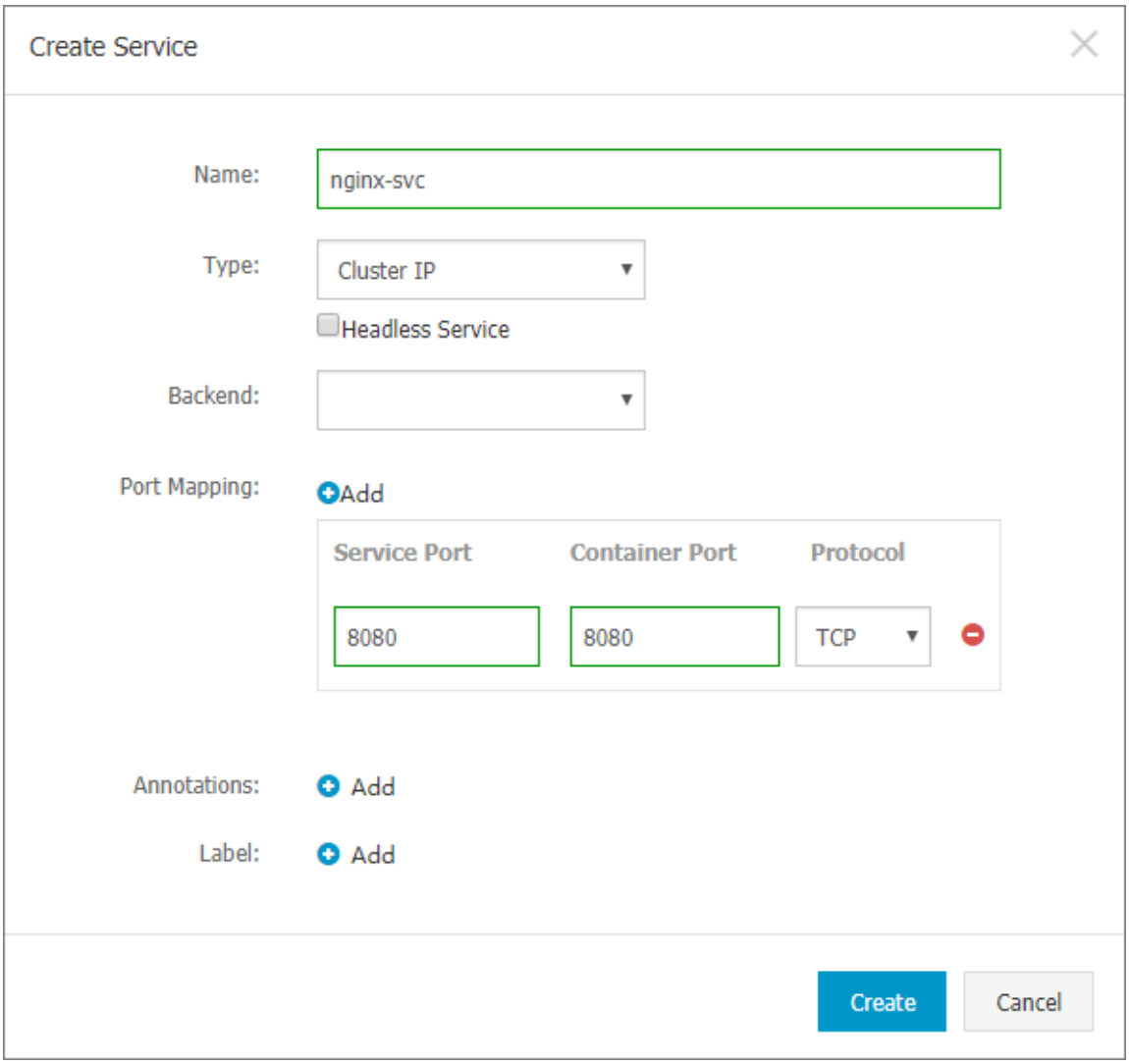
5. Set other parameters based on your needs and then click Next.
6. Configure advanced settings. Configure Access Control settings.

   You can configure how to expose the Pods and click Create. This example creates a Service of the Cluster IP type and an Ingress to build an Nginx application that is accessible to the Internet.

> **Note:**
>
> You can configure access control based on your needs:
>
> - Internal applications: For applications that run inside the cluster, you can create Services of the Cluster IP or Node Port type to enable internal communication as needed.

- **External applications: For applications that need to be exposed to the Internet, you can configure access control by using one of the following methods:**
  - **Create a Service of the Server Load Balancer type and expose your application to the Internet through the SLB instance.**
  - **Create a Service of the Cluster IP or Node Port type, create an Ingress, and expose your application to the Internet through the Ingress. For more information, see** *https://kubernetes.io/docs/concepts/services-networking/ingress/*.

a) **To create a Service, click Create in the Access Control section. Configure the Service in the dialog box that appears, and then click Create.**



Parameter	Description
Name	The service name. Default is `applicationname-svc`.

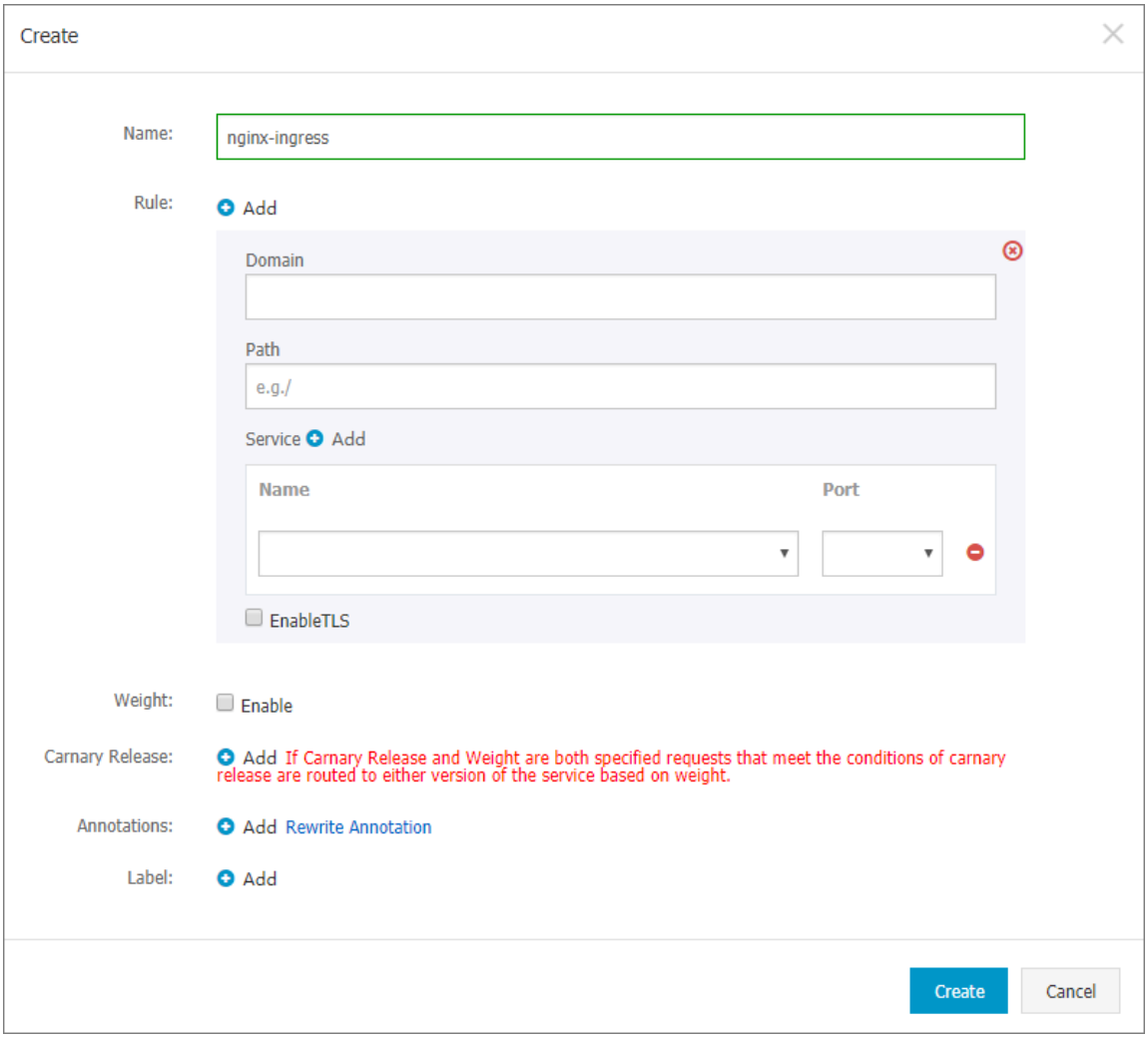
Parameter	Description
Type	Select one from the following three types. · Cluster IP: Expose the Service through an internal IP address in the cluster. When selected, the Service is only accessible within the cluster. · Node Port: Expose the Service through the IP address and static port (NodePort) on each node. The Node Port Service can route requests to a Cluster IP Service, which is automatically created by the system. You can access a Node Port Service from outside the cluster by requesting `<NodeIP>:<NodePort>`. · Server Load Balancer: Expose the Service through Server Load Balancer, which supports Internet access and internal access. Server Load Balancer can route requests to Node Port and Cluster IP Services.
Port Mapping	Set a service port and a container port. If the Type parameter is set to Node Port, you must set a node port to avoid port conflicts. TCP and UDP protocols are supported.
Annotations	Add annotations to the Service. SLB parameters are supported. For more information, see *Access services by using SLB*.
Labels	Add labels to the Service.

b) To create an Ingress, click Create in the Access Control section. Configure Ingress rules in the dialog box that appears, and then click Create. For more information about Ingress configuration, see *Ingress configurations*.

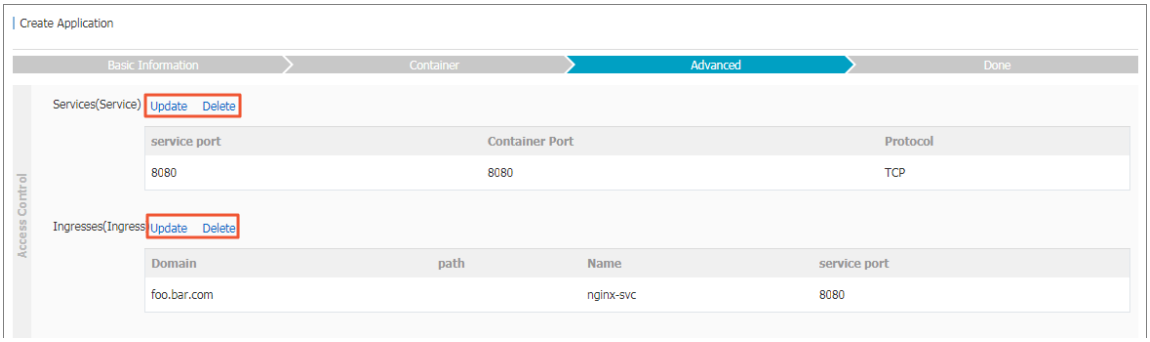
When you create an application by using an image, you can create an Ingress for one Service only. This example uses a virtual host name as the test domain.

You need to add a record to the hosts file. In actual scenarios, use a domain that has obtained an ICP filing.

```
101.37.224.146   foo.bar.com    #The IP address of the Ingress
```

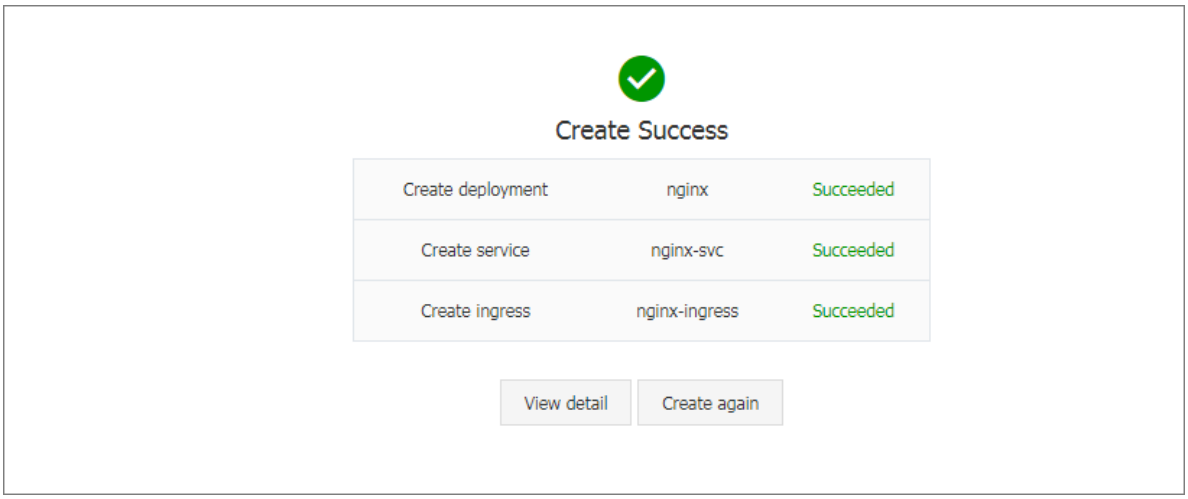


c) **You can find the newly created Service and Ingress in the Access Control section. You can click Update or Delete to make changes.**



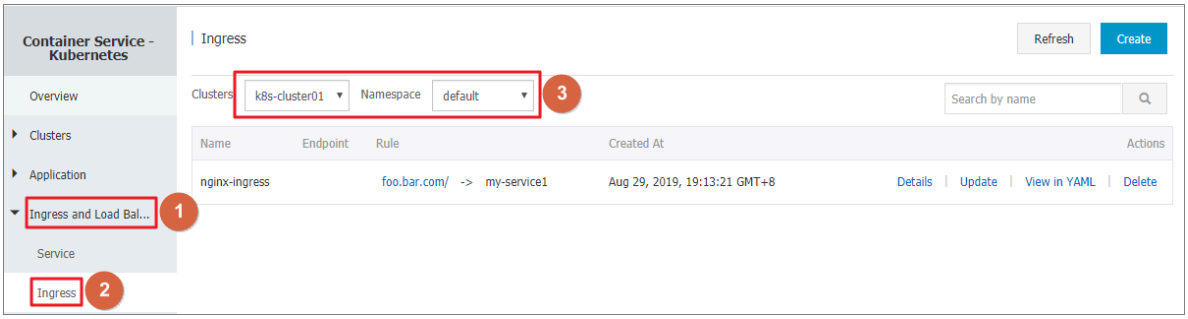
7. **Click Create.**

8. After the application is created, you are directed to the Complete page, which displays the resource objects under the application. You can click View Details to view application details.



The nginx-deployment page is displayed by default.

9. In the left-side navigation pane, choose Ingresses and Load Balancing > Ingresses. The following rule is displayed on the Ingresses page.



10. Enter the Ingress test domain into your browser and the Nginx welcome page is displayed.

3.4.5.2 Create an application from an orchestration template

Container Service provides orchestration templates that you can use to create applications quickly. You can also modify the templates based on YAML syntax to customize your applications.

Prerequisites

You have created a Kubernetes cluster. For more information, see *Create a Kubernetes cluster*.

Context

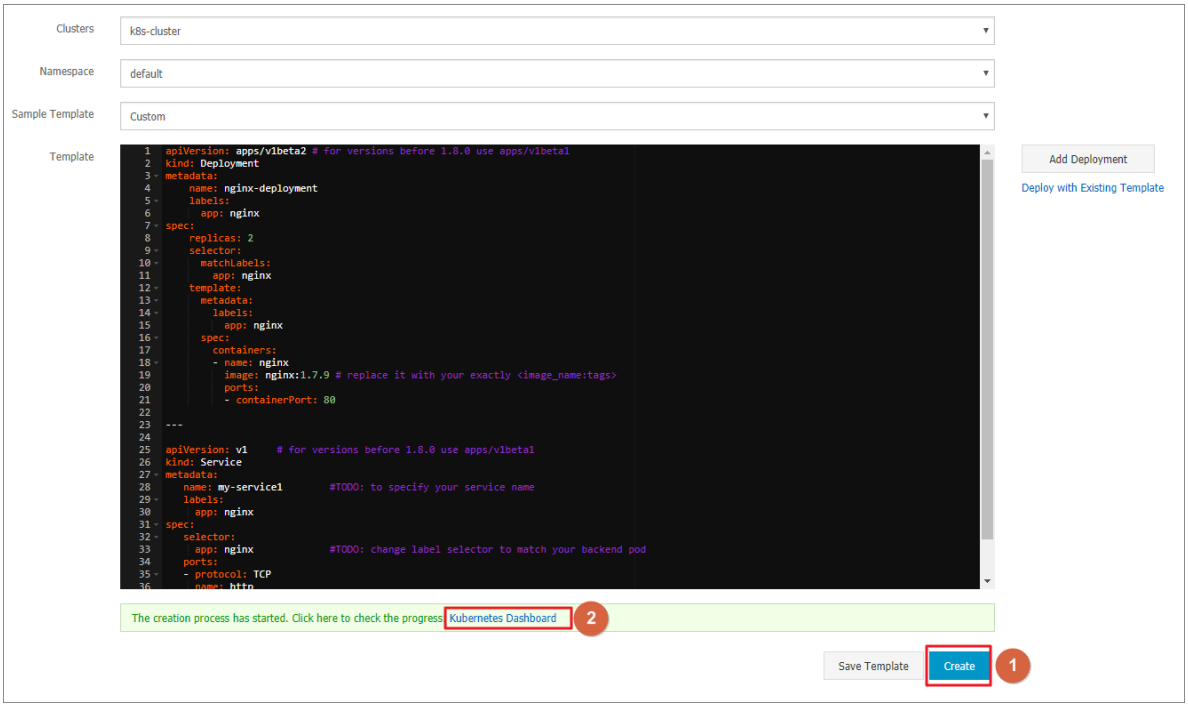
The following example demonstrates how to create an Nginx application that consists of a Deployment and a Service. The Deployment creates a Pod and the Service is then associated with the Pod.

Procedure

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Applications > Deployments to go to the Deployments page.**
3. **In the upper-right corner, click Create from Template.**
4. **Set the parameters and click Create.**
   - **Cluster: Select the cluster where the resource objects are to be deployed.**
   - **Namespace: Select the namespace that the resource objects belong to. The default namespace is default. Except for underlying computing resources such as nodes and persistent volumes, most resource objects must be divided into namespaces.**
   - **Sample Template: Container Service provides YAML templates of various resource types to help you deploy resource objects quickly. You can create a**

template based on YAML syntax to describe the resource types that you want to define.

- Add Deployment: You can quickly define a YAML template.
- Use Existing Template: You can import an existing template to the configuration page.



The following is a sample orchestration of an Nginx application. The orchestration is based on an orchestration template provided by Container Service. You can use this orchestration template to create a Deployment for an Nginx application quickly.

Note:

Container Service supports the YAML syntax and supports using the `---` symbol to separate resource objects. This allows you to create multiple resource objects in a single template.

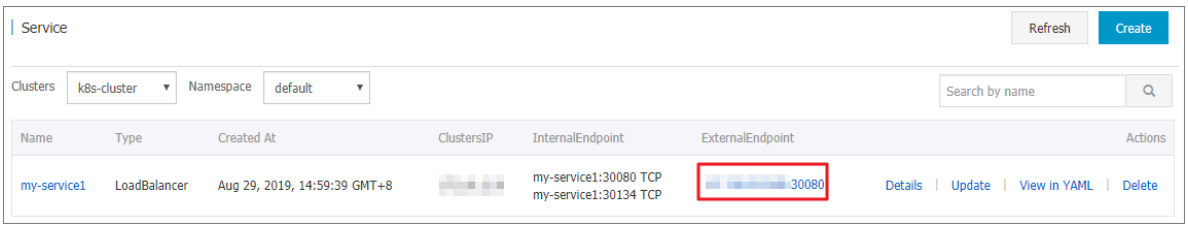
```
apiVersion: apps/v1beta2 # for versions before 1.8.0 use apps/
v1beta1
kind: Deployment
metadata:
    name: nginx-deployment
    labels:
      app: nginx
spec:
    replicas: 2
    selector:
      matchLabels:
        app: nginx
```

```
    template:
      metadata:
        labels:
          app: nginx
      spec:
        containers:
        - name: nginx
          image: nginx:1.7.9 # replace it with your exactly <
image_name:tags>
          ports:
          - containerPort: 80

---

apiVersion: v1      # for versions before 1.8.0 use apps/v1beta1
kind: Service
metadata:
   name: my-service1        #TODO: to specify your service name
   labels:
     app: nginx
spec:
   selector:
     app: nginx             #TODO: change label selector to match
your backend pod
   ports:
   - protocol: TCP
     name: http
     port: 30080            #TODO: choose an unique port on each
node to avoid port conflict
     targetPort: 80
   type: LoadBalancer        ##In the example, the type is changed
from Nodeport to LoadBalancer.
```

5. **Click Create. A message appears indicating the deployment status. You can click Kubernetes Dashboard to check the deployment progress in the dashboard.**
6. **On the Kubernetes dashboard, you can see that a Service named my-service1 is deployed and its external endpoint is displayed. Click the address under External Endpoint.**



7. **You can visit the Nginx welcome page in the browser.**

What's next

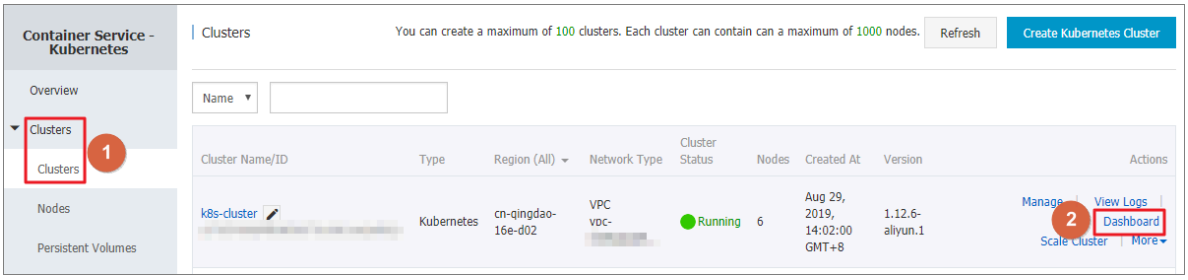
You can also choose Ingresses and Load Balancing > Services in the left-side navigation pane to view the Nginx service.

3.4.5.3 Create an application from the Kubernetes dashboard

You can create an application from the Kubernetes dashboard.

Procedure

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, click Clusters. The Clusters page appears.**
3. **Locate a cluster and click Console in the Actions column to go to the Kubernetes dashboard.**

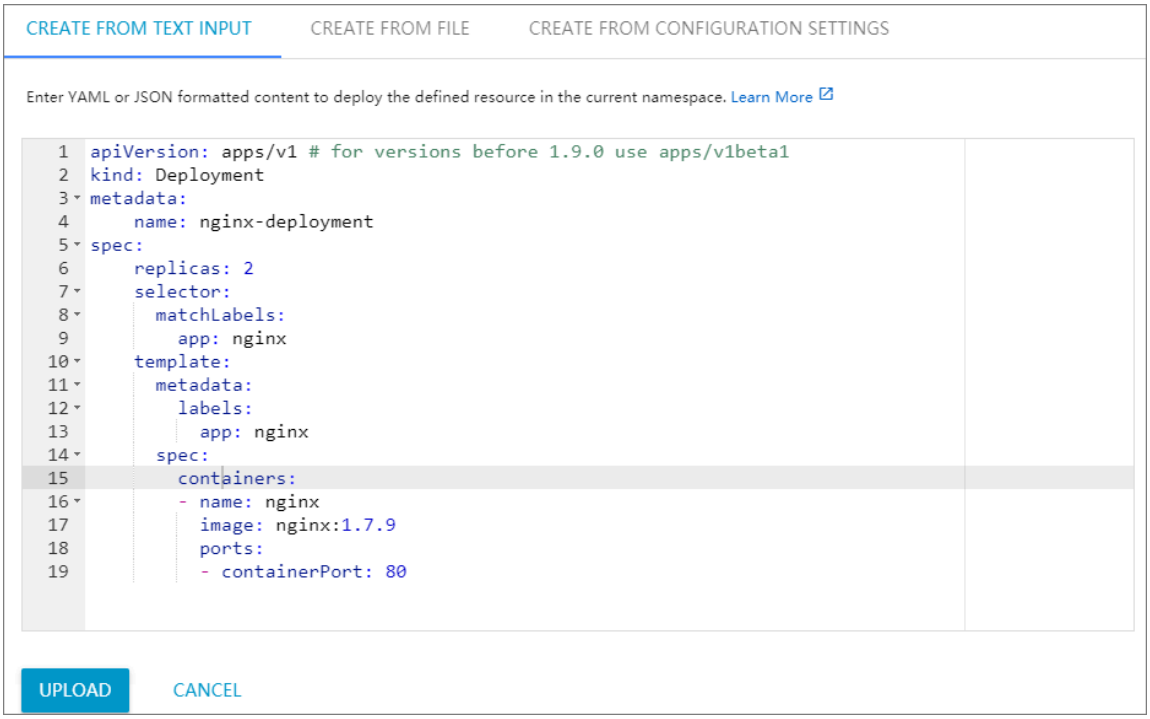


4. **In the Kubernetes dashboard, click Create in the upper-right corner.**
5. **In the dialog box that appears, set the application parameters.**

   You can use one of the following methods to create an application:

   - **Create from Text Input: Directly enter the orchestration code in YAML or JSON format to create the application. When using this method, you must**

understand the orchestration format you use well. The following figure shows an orchestration template in YAML format.

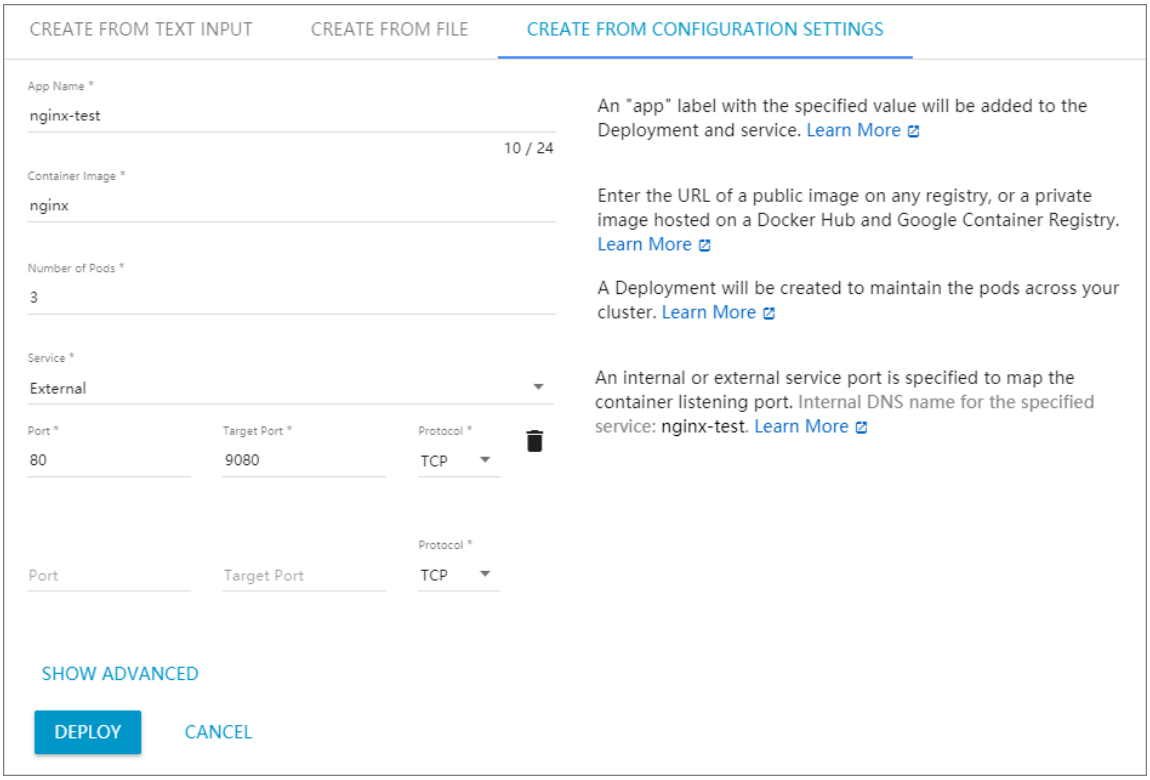


- **Create from File: Import an existing YAML or JSON configuration file to create the application.**
- **Create from Parameter Settings: Complete the following settings to create the application.**

Table 3-7: Create an application by setting parameters

Parameter	Description
App Name	The name of the application you want to create. In this example, the application name used is `nginx`.
Container Image	The URL of the image to be used. In this example, the image is Docker *NGINX*.
Number of Pods	The number of pods created for the application.
Service	Valid values: External and Internal. External indicates that the service is accessible from outside the cluster. Internal indicates that the service is accessible within the cluster.

Parameter	Description
Advanced Options	Show Advanced Options allows you to set options such as labels and environment variables. You can configure traffic to be evenly distributed across three pods by setting these options.



6. Click Deploy to deploy these containers and services.

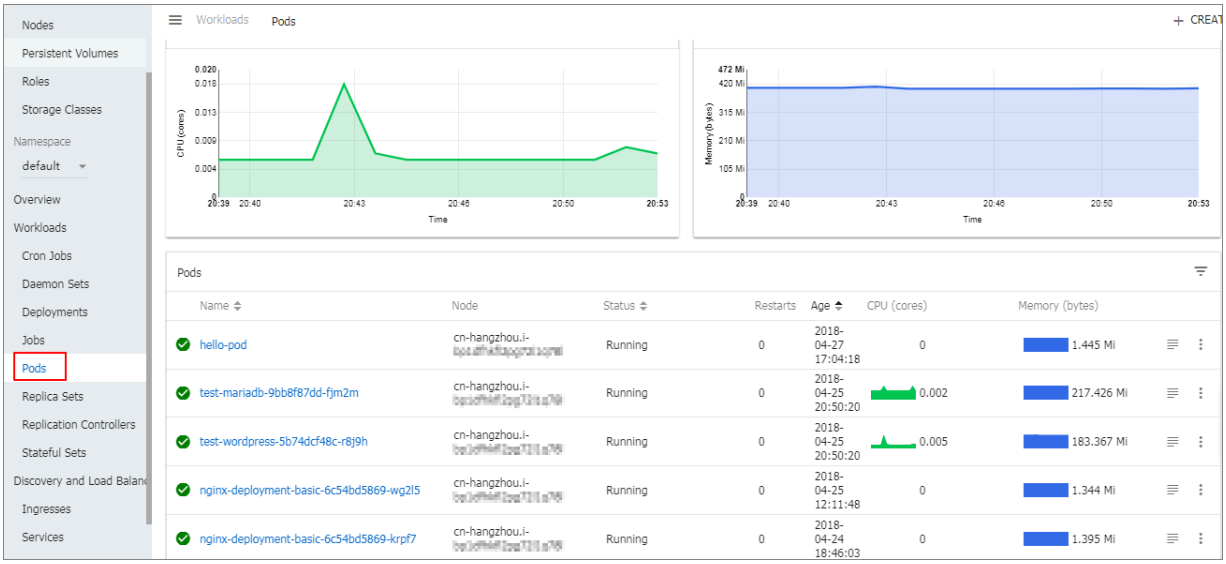
   You can also click Show Advanced Options to configure more parameters.

What's next

After clicking Deploy, you can click the left-side navigation pane of the Kubernetes dashboard to view the services or containers of the application.

In the left-side navigation pane, click Pods to check the status of each Kubernetes object by using the icons on the left. indicates the object is being deployed.

indicates the object is deployed.



3.4.5.4 Use commands to manage applications

You can use commands to create applications or view application containers.

Prerequisites

Before you use commands on your local host, you have connected to a Kubernetes cluster through kubectl. For more information, see *Connect to a Kubernetes cluster through kubectl*.

Run a command to create an application

You can use the following command to run a simple container (an NGINX Web server in this example):

```
 # kubectl run -it nginx --image=registry.aliyuncs.com/spacexnice/
netdia:latest
```

This command creates a service portal for this container. After you specify `--type=LoadBalancer`, an SLB route to the NGINX container is created.

```
# kubectl expose deployment nginx --port=80 --target-port=80 --type=
LoadBalancer
```

Run a command to view container information

Run the following command to list all running containers in the default namespace:

```
root@master # kubectl get pods
    NAME                                READY     STATUS
RESTARTS   AGE
```

```
nginx-2721357637-dvwq3                    1/1       Running   1
9h
```

3.4.5.5 Create Servcies

You can create Services for your applications through the Container Service console.

A Kubernetes Service, which is generally called a microservice, is an abstraction which defines a logical set of Pods and a policy by which to access them. The set of Pods that are accessed by the Service is usually determined by a Label Selector.

Each Pod has its own IP address. However, Pods are created and deleted dynamicall y and quickly. Using Pods to provide services externally is therefore not a high availability solution. The Service abstraction enables the decoupling between the frontend and the backend. The frontend does not need to be aware of how the backend is implemented, which leads to a loosely coupled microservices based architecture.

For more information, see *Kubernetes Service***.**

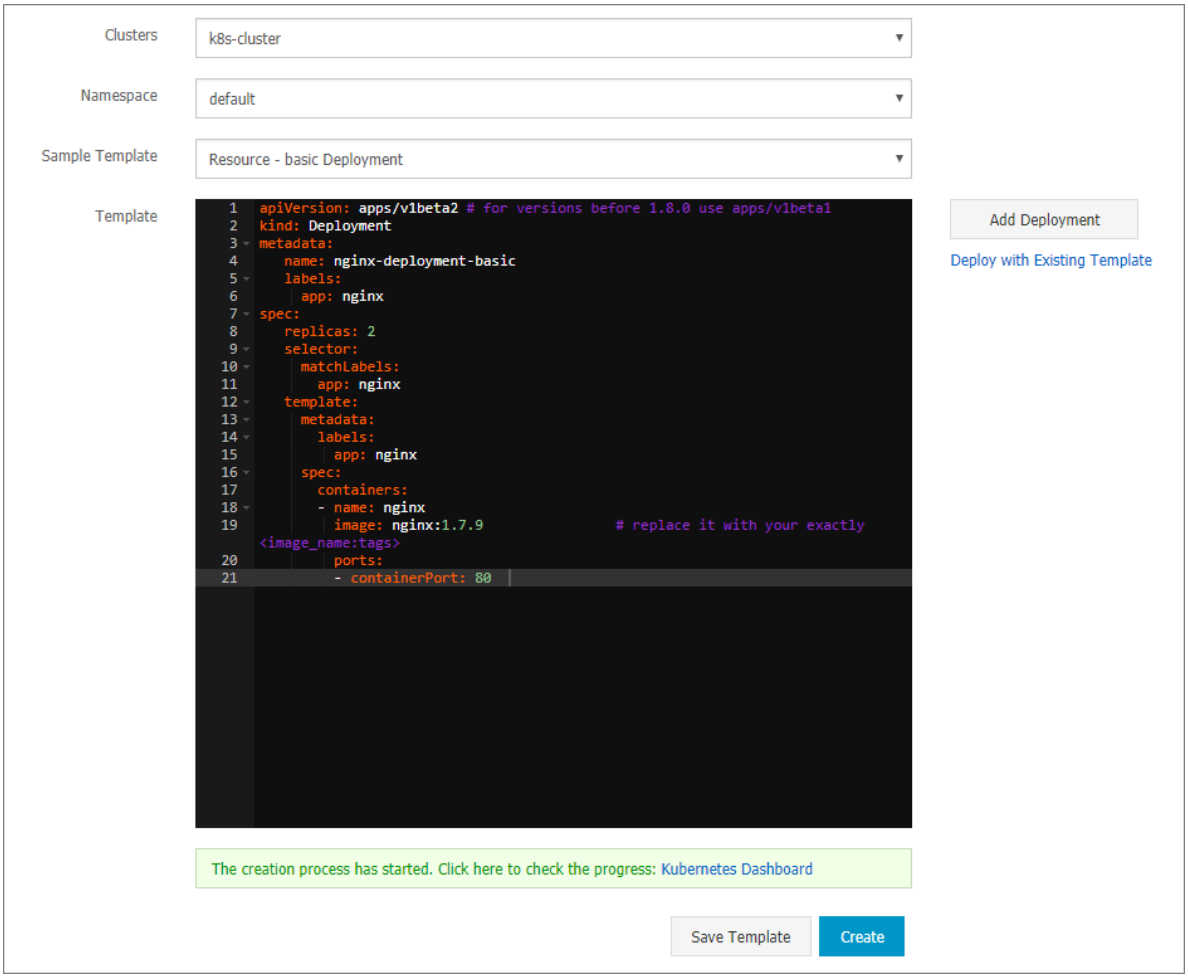
Prerequisites

You have created a Kubernetes cluster. For more information, see *Create a Kubernetes cluster***.**

Step 1: Create a Deployment

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Applications > Deployments. In the upper-right corner, click Create from Template.**

3. **Select the cluster and namespace. Then select a sample template or enter a custom template. Click Create.**



In this example, an Nginx Deployment template is used.

```
apiVersion: apps/v1beta2 # for versions before 1.8.0 use apps/
v1beta1
kind: Deployment
metadata:
   name: nginx-deployment-basic
   labels:
     app: nginx
spec:
   replicas: 2
   selector:
     matchLabels:
       app: nginx
   template:
     metadata:
       labels:
         app: nginx
     spec:
       containers:
       - name: nginx
         image: nginx:1.7.9                # replace it with your <
image_name:tags>
         ports:
```

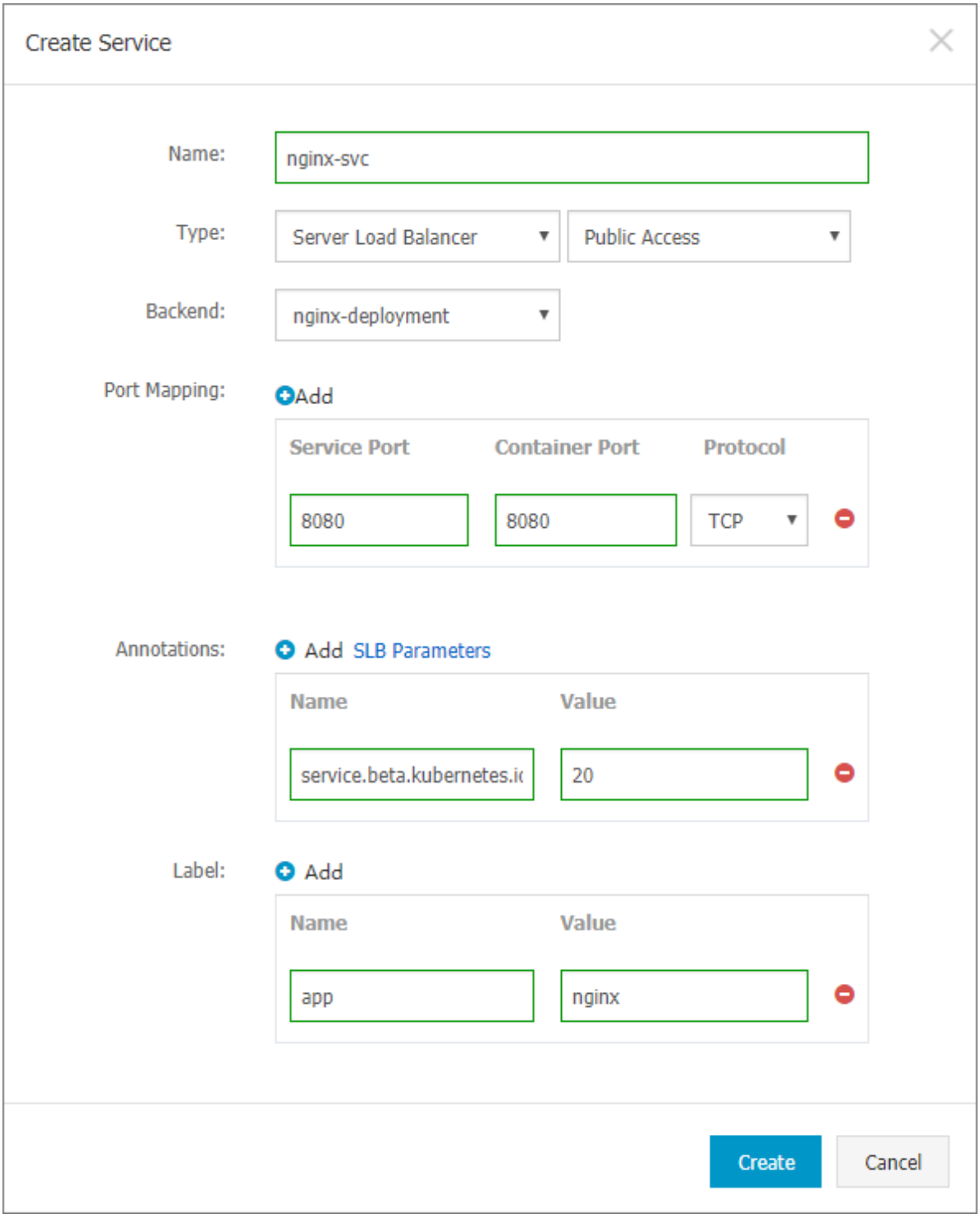
```
        - containerPort: 80                ##Expose this port in the
Service
```

4. **Click Dashboard to go to the Kubernetes dashboard and check the status of this Deployment.**

Step 2: Create a Service

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Ingresses and Load Balancing > Services to go to the Services page.**
3. **Select the cluster and namespace. Then click Create in the upper-right corner.**

4. Set the following parameters in the Create Service dialog box.

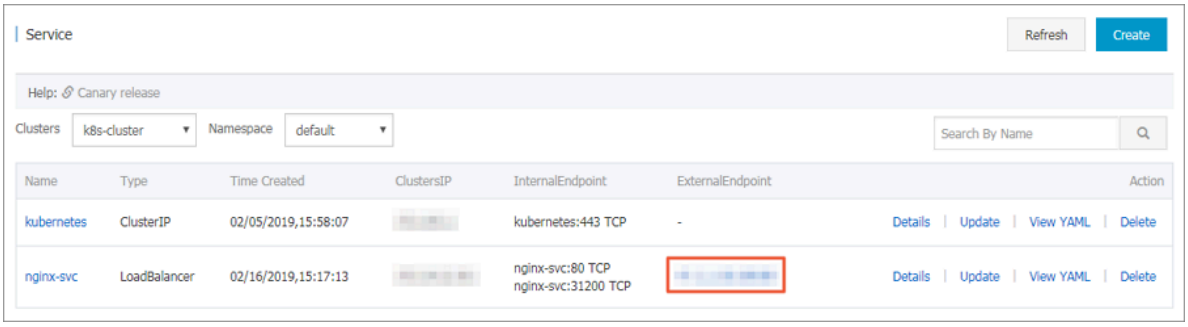


- Name: The service name. In this example, enter nginx-svc.
- Type: The service type, namely, how to expose the Service.
  - Cluster IP: Expose the Service through an internal IP address in the cluster. When selected, the Service is only accessible within the cluster. This is the default service type.

- Node Port: Expose the Service through the IP address and static port (NodePort) of the node. A Node Port Service can route requests to a Cluster IP Service, which is automatically created by the system. You can access a Node Port Service from outside the cluster by requesting `<NodeIP>:<NodePort>`.
- Server Load Balancer: Expose the Service through Server Load Balancer, which supports Internet access and internal access. Server Load Balancer can route requests to Node Port and Cluster IP Services.

- Backend: The backend object that you want to associate with the Service. In this example, select nginx-deployment-basic created from the previous step. If you do not specify a Deployment, no Endpoints object will be created. You can manually map the Service to Endpoints. For more information, see *services-without-selectors*.
- Port Mapping: Set the service port and container port. The container port must be the same as the one exposed by the backend Pod.
- Annotations: Add annotations to the Service and configure SLB parameters. For example, `service.beta.kubernetes.io/alicloud-loadbalancer-bandwidth:20` indicates that the service bandwidth is set to 20 Mbit/s. For more information, see *Access services by using SLB*.
- Labels: Add labels to the Service.

5. Click Create. You can then find the nginx-svc service on the Services page.
6. You can view basic information about the Service. You can also access its external endpoint through a browser.



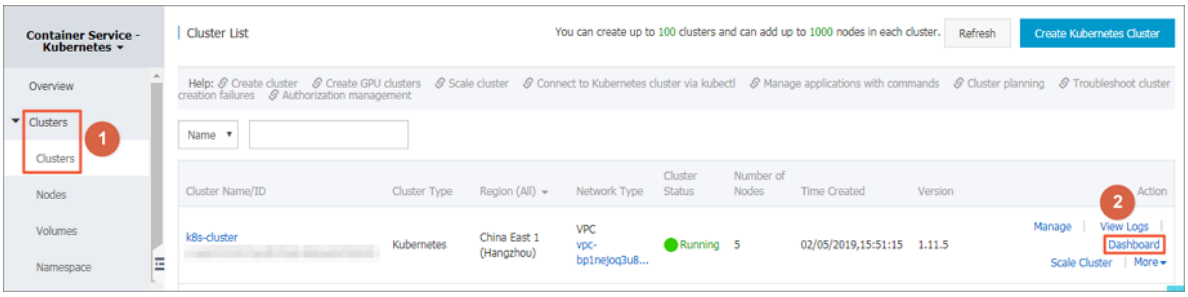
You have created a Service and associated it with a backend Deployment. You can now visit the Nginx welcome page.

3.4.5.6 Scale in or scale out a service

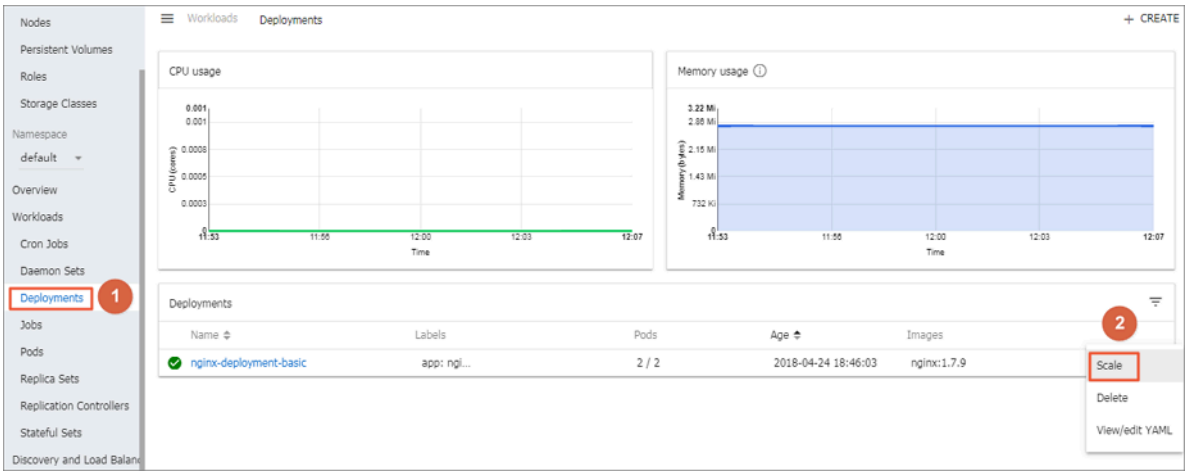
After creating an application, you can scale in or scale out the service as required.

Procedure

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, click Clusters. The Clusters page appears.**
3. **Locate a cluster and click Console in the Actions column to go to the Kubernetes dashboard.**

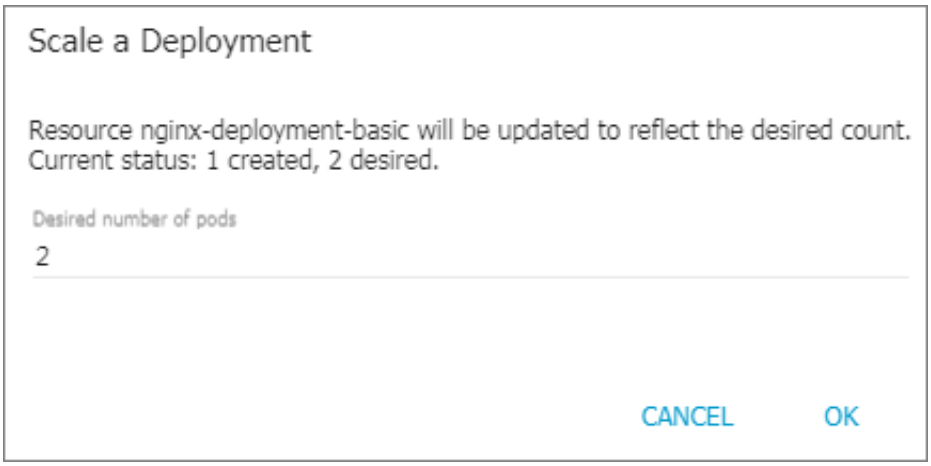


4. **In the left-side navigation pane of the Kubernetes dashboard, click Deployments to view created deployments.**
5. **Click the management icon corresponding to a deployment, and choose Scale from the shortcut menu.**



6. In the dialog box that appears, set Desired number of pods to 2 and then click OK.

   This operation adds a pod and increases the number of replicas to 2.



What's next

You can check the status of each Kubernetes object by using the icons on the left. indicates the object is being deployed. indicates the object is deployed.

After an application is deployed, you can click the name of a deployment item to view the details of the running Web service. You can view replica sets in the deployment item as well as the CPU utilization and memory usage for these replica sets. You can also click the icon to view container logs.

Note:
If no resources are displayed, wait for a few minutes and check again.

3.4.5.7 View services

This topic describes how to view the details of a service through the Container Service console.

Context

If you have configured external services during the application creation, the Kubernetes dashboard creates the external services for pre-assigned SLB instances in addition to the running containers to divert traffic to the cluster containers.

Procedure

1. *Log on to the Container Service console*.

2. **In the left-side navigation pane, choose Application > Service. The Services page appears.**
3. **Select a cluster and a namespace from the Cluster and Namespace drop-down lists. Then, click Details in the Actions column corresponding to a service.**
4. **Optional: You can also go to the Kubernetes dashboard of the cluster and click Services in the left-side navigation pane to view the services.**

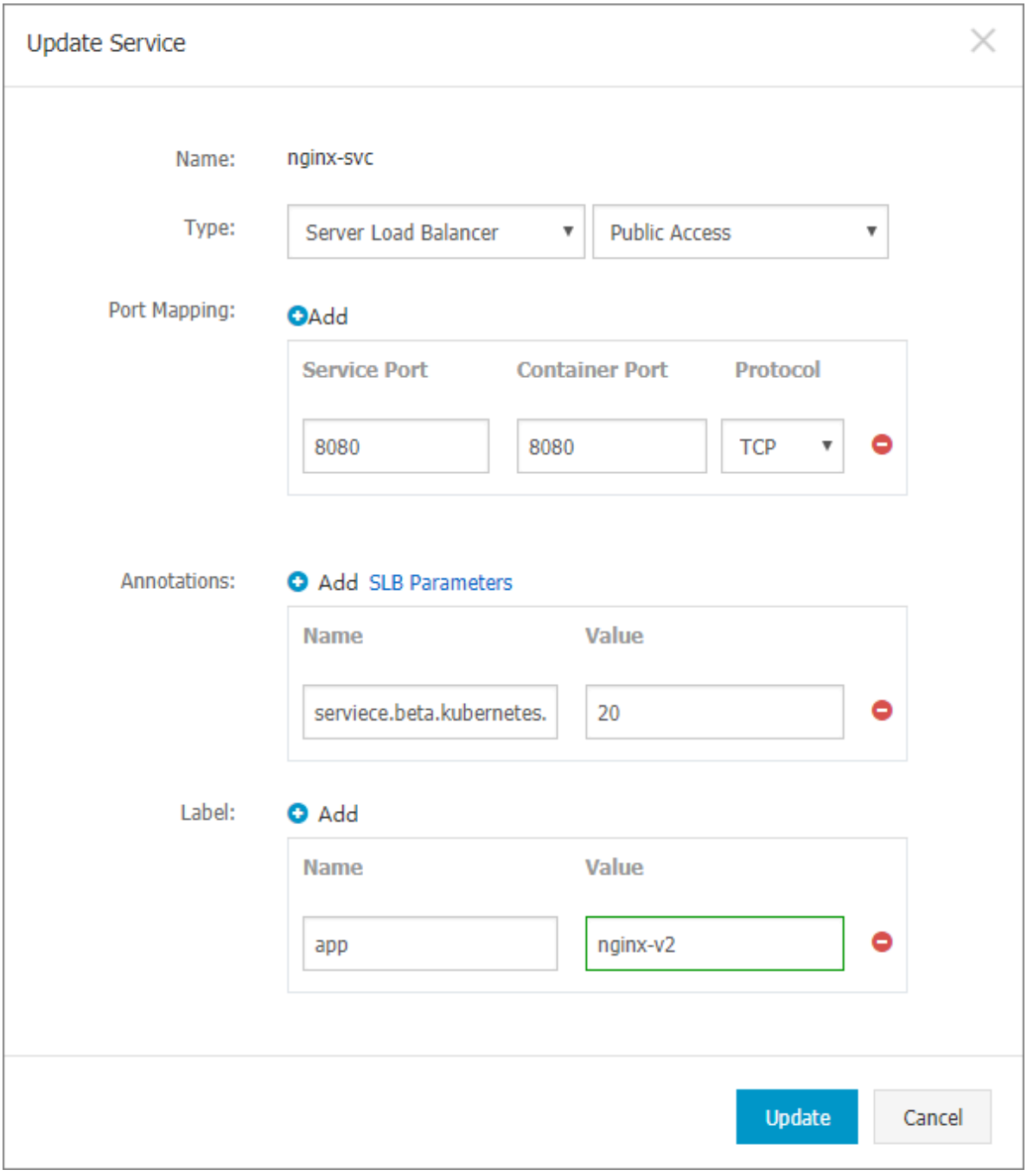
3.4.5.8 Update a service

You can update a service from the Container Service console or Kubernetes dashboard.

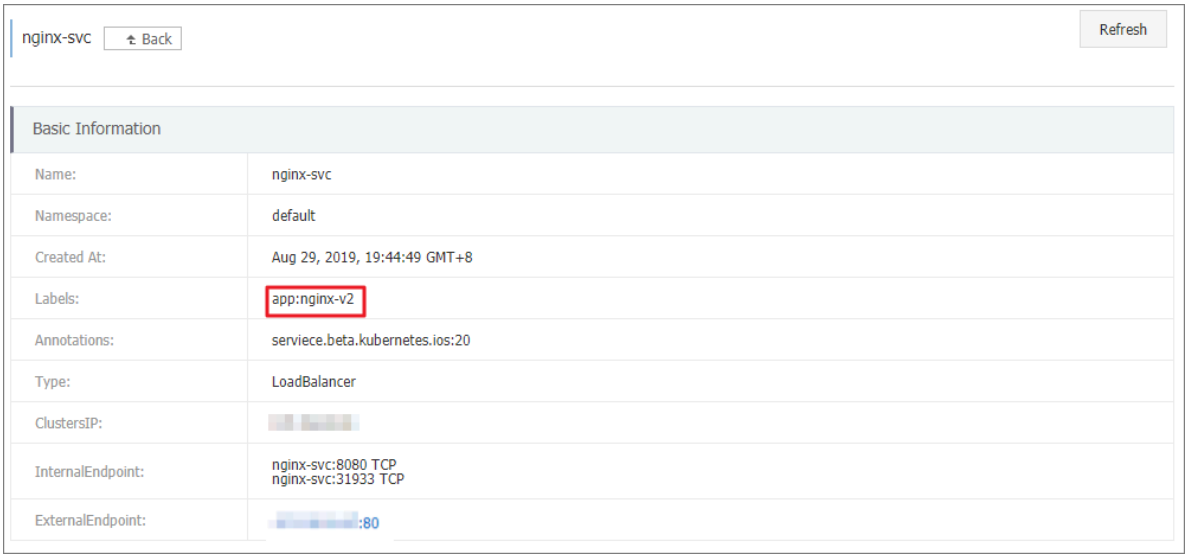
Update a service from the Container Service console

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Application > Service. The Services page appears.**
3. **Select a cluster and namespace from the Cluster and Namespace drop-down lists. Click Update in the Actions column corresponding to a service (nginx-svc in this example).**

4. **In the Update Service dialog box that appears, modify the settings and click Update.**

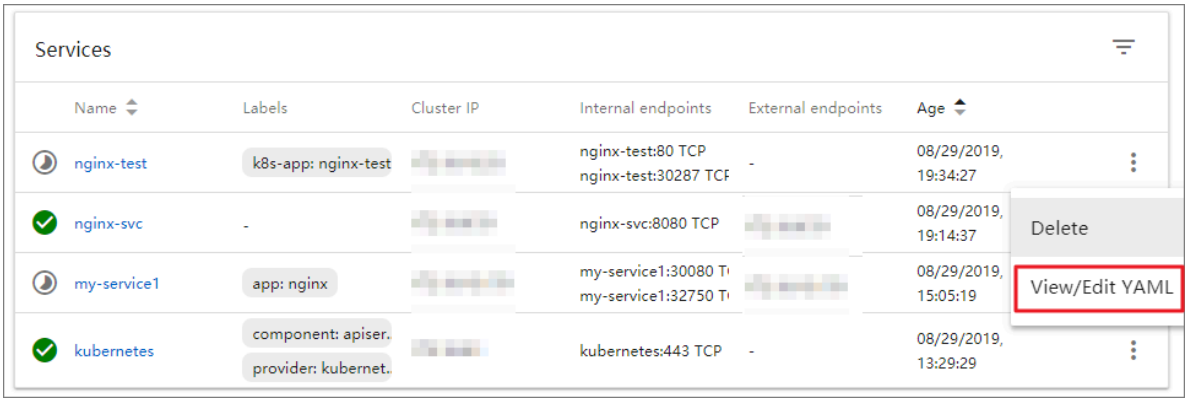


5. **Locate the service, and then click Details in the Actions column to view the changes made to the service. In this example, the service label is changed.**

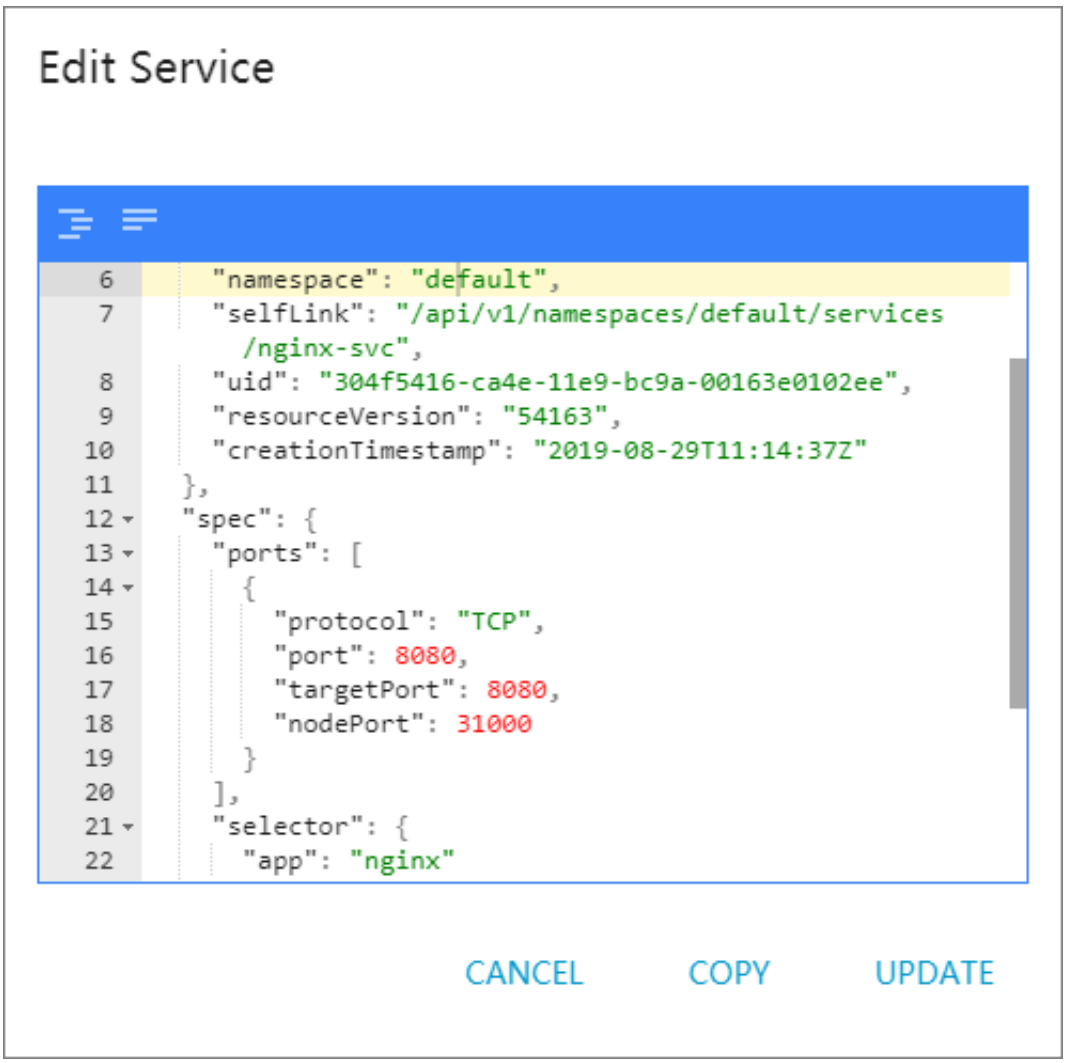


Update a service from the Kubernetes dashboard

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, click Clusters. The Clusters page appears.**
3. **Locate a cluster and click Console in the Actions column to go to the Kubernetes dashboard.**
4. **In the left-side navigation pane of the Kubernetes dashboard, select a namespace and click Services.**
5. **Click the management icon corresponding to a service and choose View/edit YAML from the shortcut menu.**



6. **In the dialog box that appears, modify the configurations. For example, set nodePort to 31000, and click Update.**



3.4.5.9 Delete a service

Prerequisites

- **You have created a Kubernetes cluster. For more information, see** *Create a Kubernetes cluster*.
- **You have created a service. For more information, see** *Create Servcies*.

Procedure

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Application > Service. The Services page appears.**

3. Select a cluster and a namespace from the Cluster and Namespace drop-down lists. Click Delete in the Actions column corresponding to a service (nginx-svc in this example).
4. In the message that appears, click Confirm. The service is then removed from the list of services.

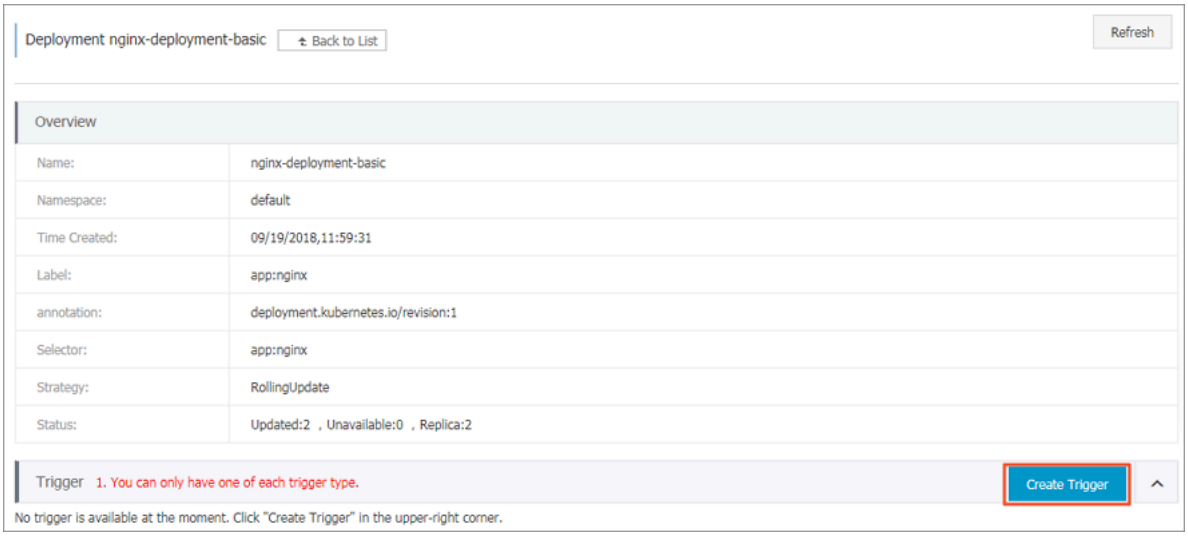
3.4.5.10 Use triggers

Prerequisites

- You have created a Kubernetes cluster. For more information, see *Create a Kubernetes cluster*.
- You have created an application that is used to create and test the trigger. This example uses an Nginx application.

Procedure

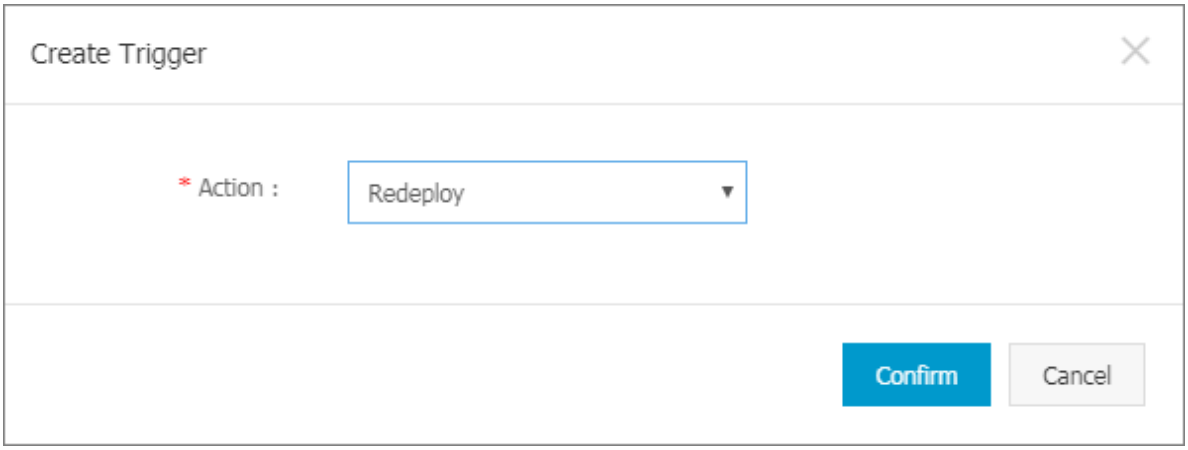
1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Applications > Deployments. Select the cluster and namespace to go to the Deployments page. Then select the Nginx application and click Details in the Actions column.
3. On the Nginx application details page, click Create Trigger in the Trigger section.



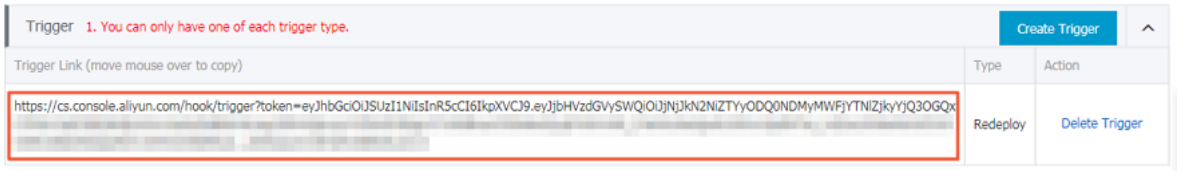
4. In the dialog box that appears, select Redeployment and click OK.

Note:

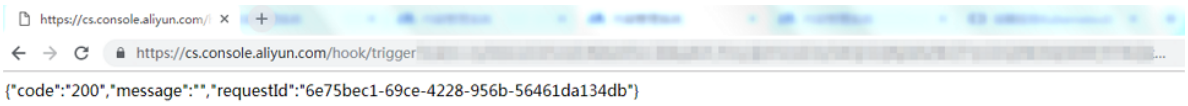
Currently, only the redeploy action is supported.



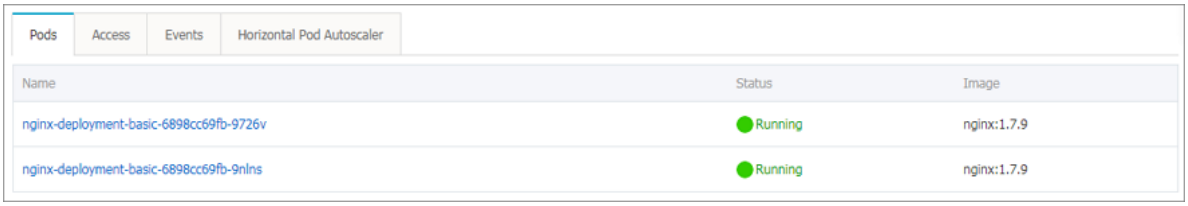
After the trigger is created, a trigger link is displayed in the Trigger section of the application details page.



5. **Copy the link and open it in the browser. A message appears, containing information such as the request ID.**



6. **Go to the Nginx application details page and a new Pod is displayed.**



After the new Pod is successfully deployed, the old Pod will be deleted.

What's next

You can call triggers through GET or POST requests from a third-party system. For example, you can run `curl` commands to call triggers.

Call the redeploy trigger as follows:

```
curl https://cs.console.aliyun.com/hook/trigger?token=xxxxxxx
```

3.4.5.11 View Pods

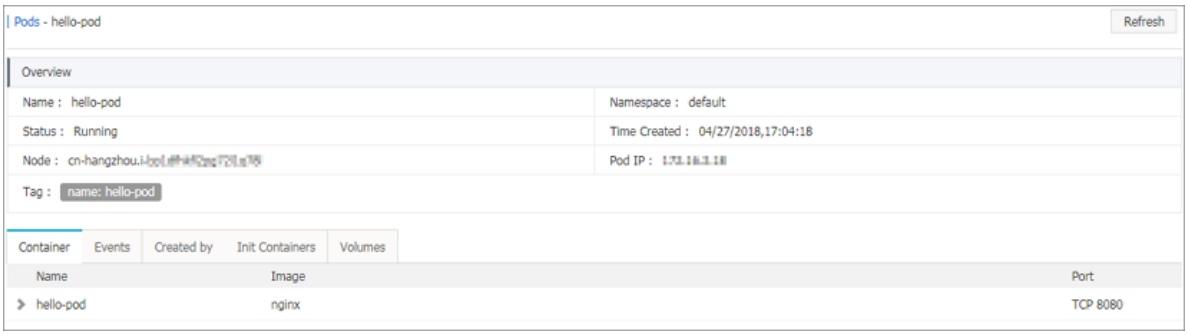
You can view Pods through the console.

View Pods on the Pods page

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Applications > Pods to go to the Pods page.**
3. **Select the cluster and namespace. Choose the Pod and click View Details.**

> **Note:**
> **You can update or delete Pods on the Pods page. We recommend that you manage Pods through Deployments if they were created through Deployments.**

4. **On the Pod details page, you can view detailed information about the selected Pod.**



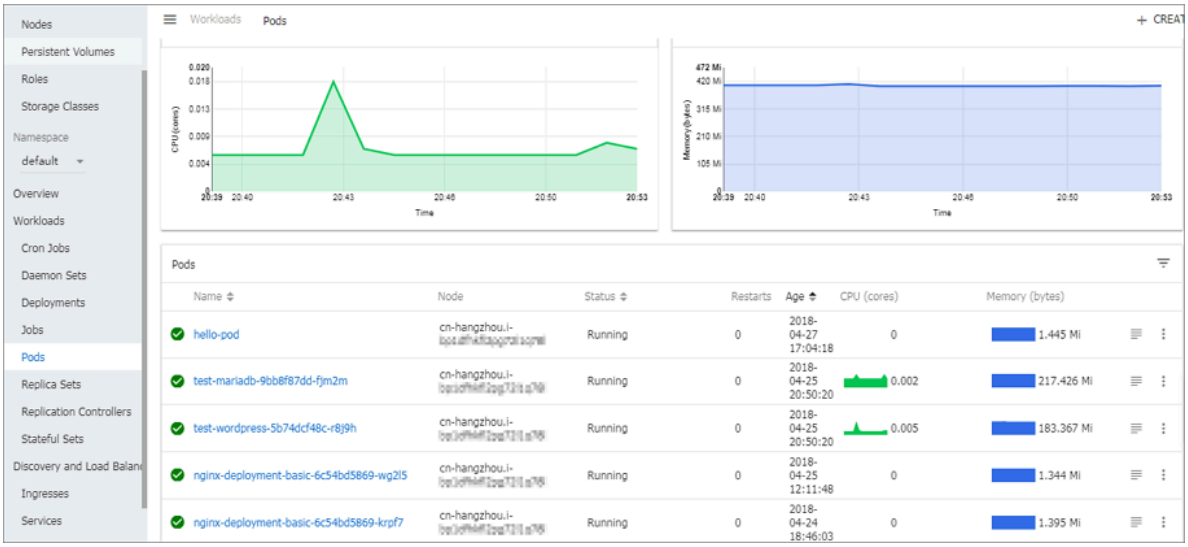
View Pods through the Kubernetes dashboard

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Clusters > Clusters to go to the Clusters page.**
3. **Choose a cluster and click Dashboard in the Actions column to go to the Kubernetes dashboard.**

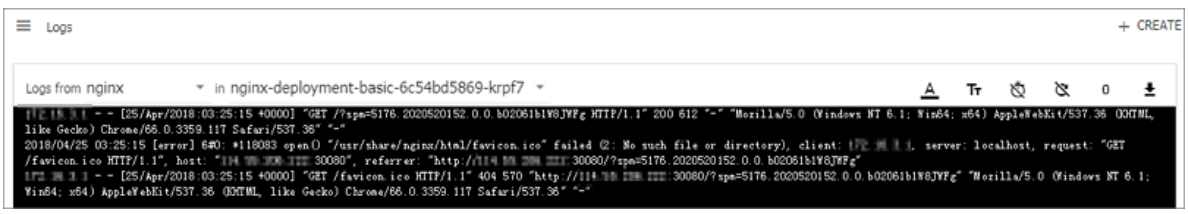
4. In the left-side navigation pane, click Pods to view the Pods in the cluster.

   You can also click Services in the left-side navigation pane and then click a Service name to view the Pods in this Service.

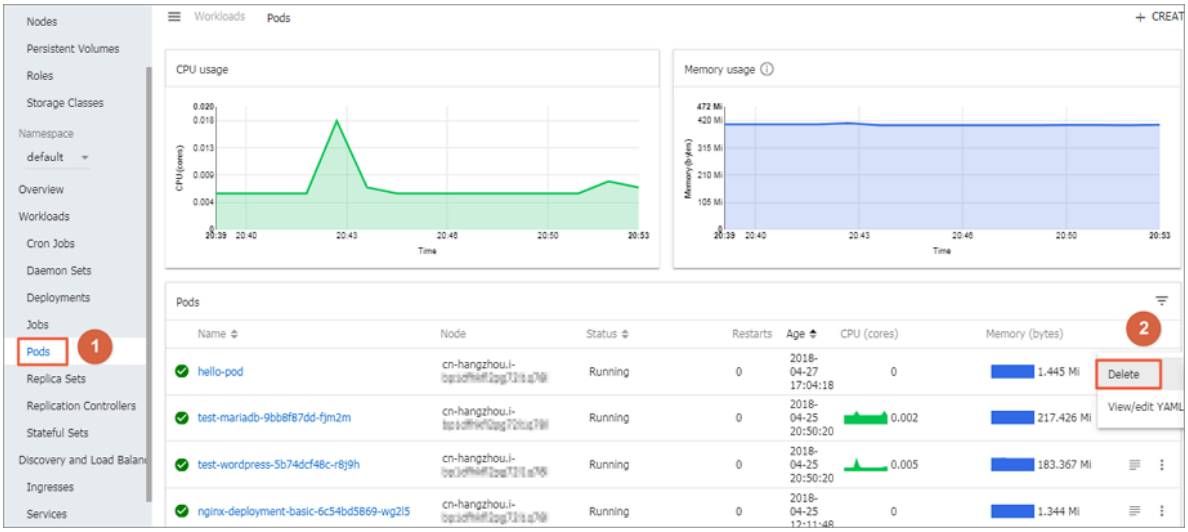
5. The icon at the left of each Pod name indicates the Pod status. indicates that the Pod is under deployment. indicates that the Pod is successfully deployed.



6. Select a Pod and click its name to view Pod details, including the CPU and memory usage.
7. Select a Pod and click at the right to view logs.



8. You can also click Delete to delete the Pod.



3.4.5.12 Schedule Pods to nodes

You can add labels to nodes and then configure `nodeSelector` to schedule Pods to specific nodes. For more information about how nodeSelector works, see *nodeselector*.

To meet business needs, you may need to deploy a service used for management and control to a master node, or deploy certain services to nodes with SSD disks. You can use the following method to schedule Pods to specific nodes based on needs.

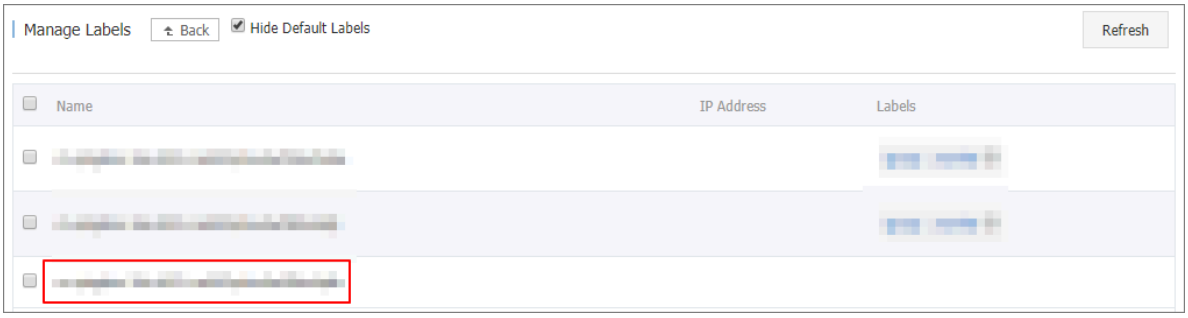
Prerequisites

You have created a Kubernetes cluster. For more information, see *Create a Kubernetes cluster*.

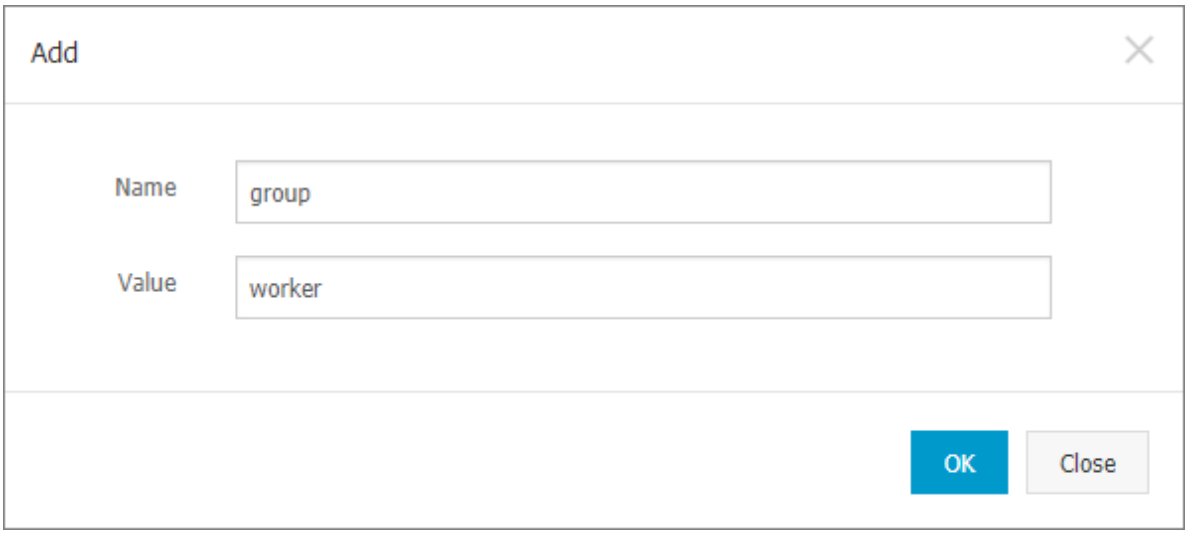
Step 1: Add labels to nodes

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Clusters > Nodes to go to the Nodes page.
3. Select the cluster and click Manage Labels in the upper-right corner.

4. **Select one or multiple nodes and then click Add Label. In this example, select a worker node.**



5. **In the dialog box that appears, enter the name and value of the label and then click OK.**



On the Manage Labels page, you can find the `group:worker` label next to the selected node.

You can also use the following command to add labels to nodes: `kubectl label nodes <node-name> <label-key>=<label-value>`.

Step 2: Schedule Pods to specific nodes

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Applications > Deployments to go to the Deployments page.**
3. **In the upper-right corner, click Create from Template.**

4. **Configure the template to create a Pod and schedule the Pod to the specific node. After the configuration is complete, click Create.**
   - **Cluster: Select the cluster where the resource object is deployed.**
   - **Namespace: Select the namespace that the resource object belongs to. In this example, use the default namespace.**
   - **Sample Template: Select Custom in this example.**

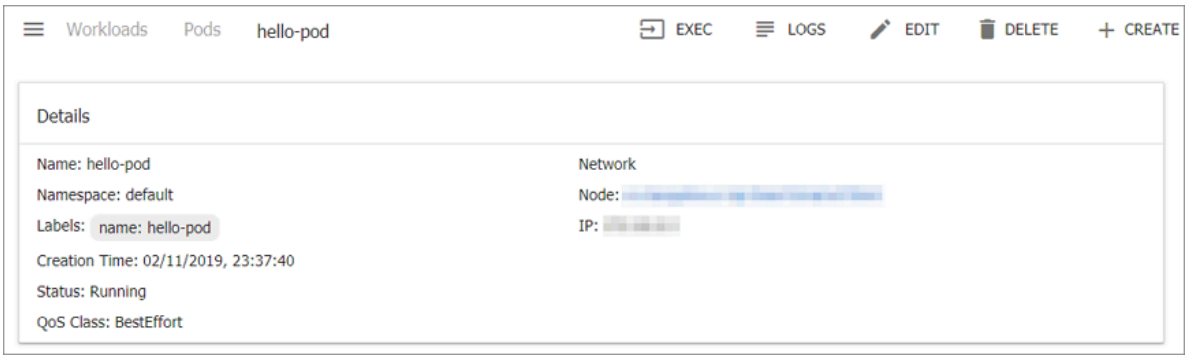
   Enter the following orchestration template:

```
 apiVersion: v1
     kind: Pod
     metadata:
     labels:
     name: hello-pod
     name: hello-pod
     spec:
     containers:
     - image: nginx
     imagePullPolicy: IfNotPresent
     name: hello-pod
     ports:
     - containerPort: 8080
     protocol: TCP
     resources: {}
     securityContext:
     capabilities: {}
     privileged: false
     terminationMessagePath: /dev/termination-log
     dnsPolicy: ClusterFirst
     restartPolicy: Always
     nodeSelector:
     group: worker  ##Note that this value must be the same as the
node label created from the previous step.
     status: {}
```

5. **Click Create and a message appears indicating the deployment status. After the deployment is complete, click Dashboard to view the Pod status on the Kubernetes dashboard.**

6. **Click the name of the Pod to view its details.**

   You can find its label and the ID of the node that the Pod is scheduled to. The following figure indicates that the Pod is now scheduled to the node with the `group:worker` label.



3.4.6 SLB and Ingress

3.4.6.1 Overview

Container Service allows you to flexibly manage load balancing and customize load balancing policies for Kubernetes clusters. Kubernetes clusters provide you with a variety of methods to access containerized applications. They also allow you to use SLB or Ingress to access internal services and implement load balancing.

3.4.6.2 Access services by using SLB

You can access services by using Apsara Stack Server Load Balancer (SLB).

Use command lines

1. **Create an Nginx application by using command lines.**

```
root@master # kubectl run nginx --image=registry.aliyuncs.com/acs/
netdia:latest
       root@master # kubectl get po
       NAME                                READY     STATUS
RESTARTS   AGE
       nginx-2721357637-dvwq3              1/1       Running   1
         6s
```

2. **Create a Service of the Server Load Balancer type for the Nginx application. Set `type=LoadBalancer` to expose the Nginx application to the Internet.**

```
root@master # kubectl expose deployment nginx --port=80 --target-
port=80 --type=LoadBalancer
       root@master # kubectl get svc
       NAME                  CLUSTER-IP      EXTERNAL-IP      PORT(S
)                       AGE
```

```
        nginx                    172.19.XX.XX     101.37.XX.XX      80:
31891/TCP                    4s
```

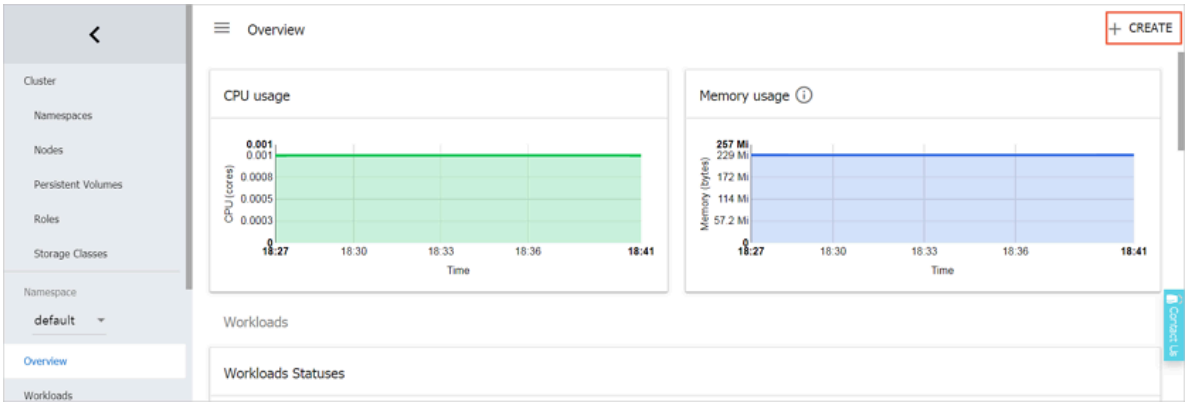
3. **Open the link `http://101.37.192.20` in a browser to access the Nginx service.**

Use the Kubernetes dashboard

1. **In the local environment, save the following YAML code to the `nginx-svc.yml` file.**

```
apiVersion: v1
kind: Service
metadata:
  labels:
    run: nginx
  name: http-svc
  namespace: default
spec:
  ports:
  - port: 80
    protocol: TCP
    targetPort: 80
  selector:
    run: nginx
  type: LoadBalancer
```

2. *Log on to the Container Service console*.
3. **Select the target cluster and click Dashboard in the Actions column to go to the Kubernetes dashboard.**
4. **Click Create in the upper-right corner to create an application.**

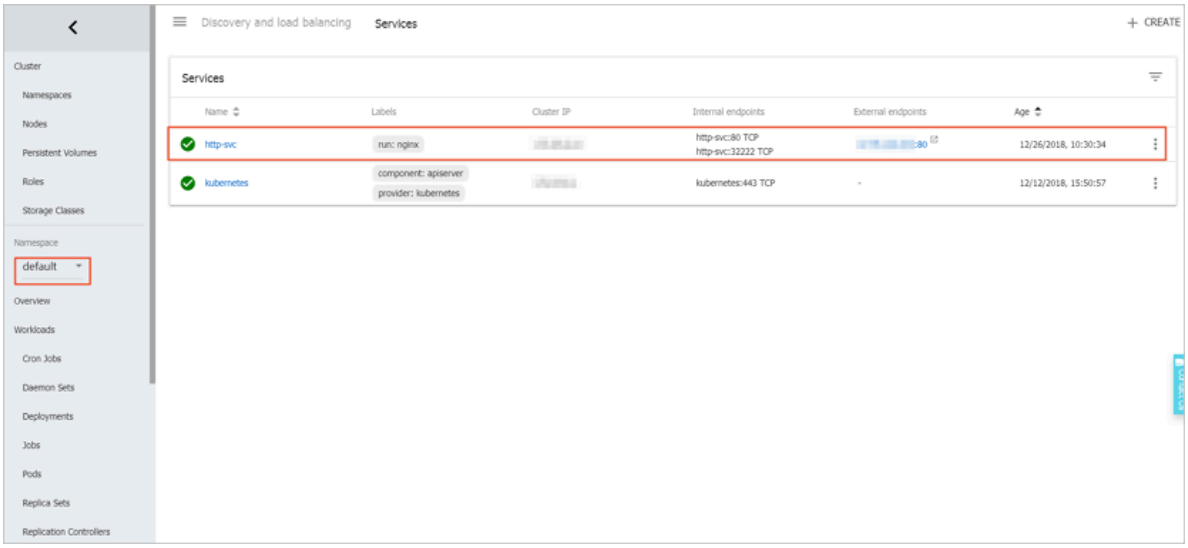


5. **Click the CREATE FROM FILE tab. Choose the *`nginx-svc.yml`* file you just saved.**
6. **Click Upload.**

   This creates a Service of the Server Load Balancer type and associates the Service with the Nginx application. The service is named `http-svc`.

7. **On the Kubernetes dashboard, select the default namespace and then click Services.**

   You can find the newly created `http-svc` Service and its external endpoint `http://114.55.XX.XX:80`.



8. **To access the service, open the link in a browser.**

More information

Apsara Stack Server Load Balancer supports a variety of parameters, such as health check, billing method, and load balancing type. For more information, see *Table 3-8: SLB instance parameters*.

Annotations

You can add annotations to use the features provided by Server Load Balancer.

Use existing internal SLB instances

You need to add two annotations as follows. Note that you must replace "yourloadbalancer-id" with your own SLB instance ID.

```
apiVersion: v1
kind: Service
metadata:
  annotations:
    service.beta.kubernetes.io/alicloud-loadbalancer-address-type:
intranet
    service.beta.kubernetes.io/alicloud-loadbalancer-id: your-
loadbalancer-id
  labels:
    run: nginx
  name: nginx
  namespace: default
spec:
  ports:
```

```
  - name: web
    port: 80
    protocol: TCP
    targetPort: 80
  selector:
    run: nginx
  sessionAffinity: None
  type: LoadBalancer
```

Save the above content as slb.svc and run the following command: `kubectl apply -f slb.svc`.

Create an HTTPS-based Service of the Server Load Balancer type

Create a certificate in the Apsara Stack console and copy the cert-id. Then create an HTTPS-based Service of the Server Load Balancer type as follows:

```
apiVersion: v1
kind: Service
metadata:
  annotations:
    service.beta.kubernetes.io/alicloud-loadbalancer-cert-id: your-
cert-id
    service.beta.kubernetes.io/alicloud-loadbalancer-protocol-port: "
https:443"
  labels:
    run: nginx
  name: nginx
  namespace: default
spec:
  ports:
  - name: web
    port: 443
    protocol: TCP
    targetPort: 443
  selector:
    run: nginx
  sessionAffinity: None
  type: LoadBalancer
```

Note:

Annotations are case sensitive.

Table 3-8: SLB instance parameters

Annotation	Description	Default value
service.beta.kubernetes.io/alicloud-loadbalancer-protocol-port	Separate multiple values with commas (,). For example, https:443, http:80.	None
service.beta.kubernetes.io/alicloud-loadbalancer-address-type	Valid values: internet and intranet.	internet

Annotation	Description	Default value
service.beta.kubernetes.io/alicloud-loadbalancer-slb-network-type	The network type of the SLB instance. Valid values: classic and vpc.	classic
service.beta.kubernetes.io/alicloud-loadbalancer-charge-type	Valid values: paybytraffic and paybybandwidth.	paybybandwidth
service.beta.kubernetes.io/alicloud-loadbalancer-id	The ID of the SLB instance. You can use loadbalancer-id to specify an existing SLB instance and the existing listener will be overwritten. The SLB instance will not be deleted if the Service is deleted.	None
service.beta.kubernetes.io/alicloud-loadbalancer-backend-label	Use labels to specify which nodes are mounted to the backend of the SLB instance.	None
service.beta.kubernetes.io/alicloud-loadbalancer-region	The region where the SLB instance is located.	None
service.beta.kubernetes.io/alicloud-loadbalancer-bandwidth	The bandwidth of the SLB instance.	50
service.beta.kubernetes.io/alicloud-loadbalancer-cert-id	The certificate ID. You need to upload the certificate first.	“”
service.beta.kubernetes.io/alicloud-loadbalancer-health-check-flag	Valid values: on and off.	Default is off. When TCP is used, do not modify this parameter. Health check is enabled by default when TCP is used.
service.beta.kubernetes.io/alicloud-loadbalancer-health-check-type	See the CreateLoadBalancerTCPListener section in the *Server Load Balancer Developer Guide*.	None

Annotation	Description	Default value
service.beta.kubernetes.io/alicloud-loadbalancer-health-check-uri	**See the CreateLoadBalancerTCPListener section in the** *Server Load Balancer Developer Guide* **.**	**None**
service.beta.kubernetes.io/alicloud-loadbalancer-health-check-connect-port	**See the CreateLoadBalancerTCPListener section in the** *Server Load Balancer Developer Guide* **.**	**None**
service.beta.kubernetes.io/alicloud-loadbalancer-healthy-threshold	**See the CreateLoadBalancerTCPListener section in the** *Server Load Balancer Developer Guide* **.**	**None**
service.beta.kubernetes.io/alicloud-loadbalancer-unhealthy-threshold	**See the CreateLoadBalancerTCPListener section in the** *Server Load Balancer Developer Guide* **.**	**None**
service.beta.kubernetes.io/alicloud-loadbalancer-health-check-interval	**See the CreateLoadBalancerTCPListener section in the** *Server Load Balancer Developer Guide* **.**	**None**
service.beta.kubernetes.io/alicloud-loadbalancer-health-check-connect-timeout	**See the CreateLoadBalancerTCPListener section in the** *Server Load Balancer Developer Guide* **.**	**None**
service.beta.kubernetes.io/alicloud-loadbalancer-health-check-timeout	**See the CreateLoadBalancerTCPListener section in the** *Server Load Balancer Developer Guide* **.**	**None**

3.4.6.3 Configure Ingress monitoring

You can enable the default VTS module of Ingress to view Ingress monitoring data.

Use command lines

1. **Modify the Ingress ConfigMap configuration by adding the following configuration item:** `enable-vts-status: "true"`**.**

```
root@master # kubectl edit configmap nginx-configuration -n kube-
system
```

```
      configmap "nginx-configuration" edited
```

After the ConfigMap configuration is modified, Ingress ConfigMap details are as follows:

```
apiVersion: v1
data:
  enable-vts-status: "true" # Enable the VTS module.
  proxy-body-size: 20m
kind: ConfigMap
metadata:
  annotations:
    kubectl.kubernetes.io/last-applied-configuration: |
      {"apiVersion":"v1","data":{"proxy-body-size":"20m"},"kind":"
ConfigMap","metadata":{"annotations":{},"labels":{"app":"ingress-
nginx"},"name":"nginx-configuration","namespace":"kube-system"}}
  creationTimestamp: 2018-03-20T07:10:18Z
  labels:
    app: ingress-nginx
  name: nginx-configuration
  namespace: kube-system
  selfLink: /api/v1/namespaces/kube-system/configmaps/nginx-
configuration
```

2. **Verify that the VTS module is enabled for Ingress NGINX.**

```
root@master # kubectl get pods --selector=app=ingress-nginx -n kube-
system
NAME                                        READY     STATUS
RESTARTS   AGE
nginx-ingress-controller-79877595c8-78gq8   1/1       Running   0
       1h
root@master # kubectl exec -it nginx-ingress-controller-79877595c8-
78gq8 -n kube-system -- cat /etc/nginx/nginx.conf | grep vhost_traf
fic_status_display
         vhost_traffic_status_display;
         vhost_traffic_status_display_format html;
```

3. **Access the Ingress NGINX console from a local device.**

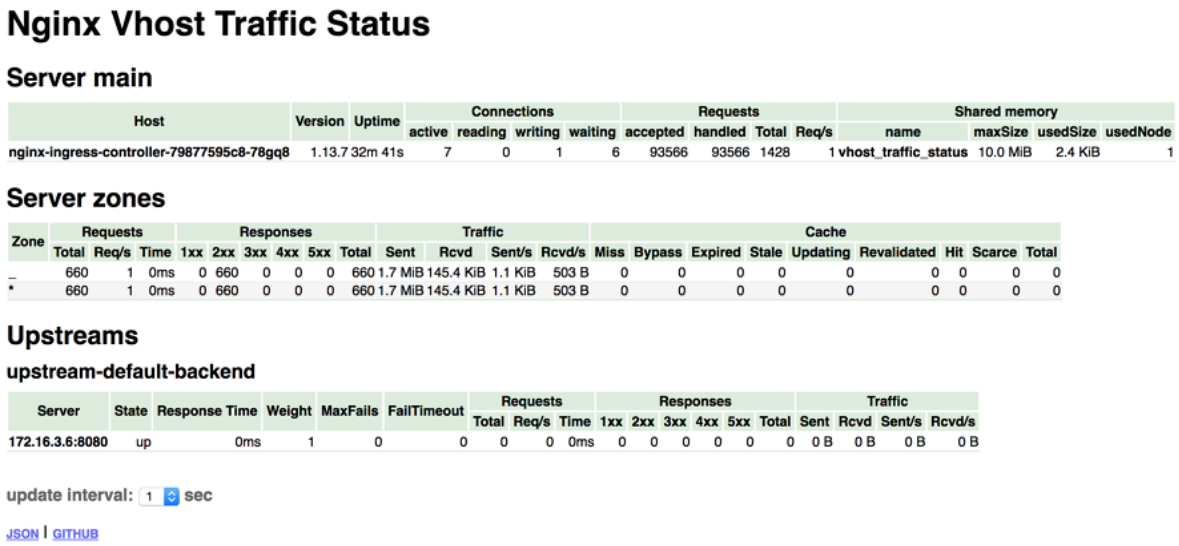
Note:

By default, the VTS port is not exposed to ensure security. In this example, port forwarding is used for access.

```
root@master # kubectl port-forward nginx-ingress-controller-
79877595c8-78gq8 -n kube-system 18080
Forwarding from 127.0.0.1:18080 -> 18080
```

```
Handling connection for 18080
```

4. **Use `http://localhost:18080/nginx_status` to access the VTS monitoring console.**



Nginx Vhost Traffic Status

Server main

Host	Version	Uptime	Connections				Requests				Shared memory			
			active	reading	writing	waiting	accepted	handled	Total	Req/s	name	maxSize	usedSize	usedNode
nginx-ingress-controller-79877595c8-78gq8	1.13.7	32m 41s	7	0	1	6	93566	93566	1428	1	vhost_traffic_status	10.0 MiB	2.4 KiB	1

Server zones

Zone	Requests			Responses					Traffic				Cache								
	Total	Req/s	Time	1xx	2xx	3xx	4xx	5xx	Total	Sent	Rcvd	Sent/s	Rcvd/s	Miss	Bypass	Expired	Stale	Updating	Revalidated	Hit	Scarce	Total
_	660	1	0ms	0	660	0	0	0	660	1.7 MiB	145.4 KiB	1.1 KiB	503 B	0	0	0	0	0	0	0	0	0
*	660	1	0ms	0	660	0	0	0	660	1.7 MiB	145.4 KiB	1.1 KiB	503 B	0	0	0	0	0	0	0	0	0

Upstreams

upstream-default-backend

Server	State	Response Time	Weight	MaxFails	FailTimeout	Requests			Responses						Traffic			
						Total	Req/s	Time	1xx	2xx	3xx	4xx	5xx	Total	Sent	Rcvd	Sent/s	Rcvd/s
172.16.3.6:8080	up	0ms	1	0	0	0	0	0ms	0	0	0	0	0	0	0 B	0 B	0 B	0 B

update interval: 1 sec

JSON | GITHUB

Use the Kubernetes dashboard

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, click Clusters. On the Clusters page, locate a cluster and click Console in the Actions column to go to the Kubernetes dashboard.**
3. **In the left-side navigation pane, select kube-system from the Namespace drop-down list. Click Config Maps. On the Config Maps page, click the vertical dots corresponding to nginx-configuration and then choose View/edit YAML from the shortcut menu. Edit the config map by adding the following configuration item: `enable-vts-status: "true"`.**

   The content of the saved Ingress ConfigMap is as follows:

```
{
  "kind": "ConfigMap",
  "apiVersion": "v1",
  "metadata": {
    "name": "nginx-configuration",
    "namespace": "kube-system",
    "selfLink": "/api/v1/namespaces/kube-system/configmaps/nginx-
configuration",
    "creationTimestamp": "2018-03-20T07:10:18Z",
    "labels": {
      "app": "ingress-nginx"
    },
    "annotations": {
      "kubectl.kubernetes.io/last-applied-configuration": "{\"
apiVersion\":\"v1\",\"data\":{\"proxy-body-size\":\"20m\"},\"kind\":
\"ConfigMap\",\"metadata\":{\"annotations\":{},\"labels\":{\"app\":
```

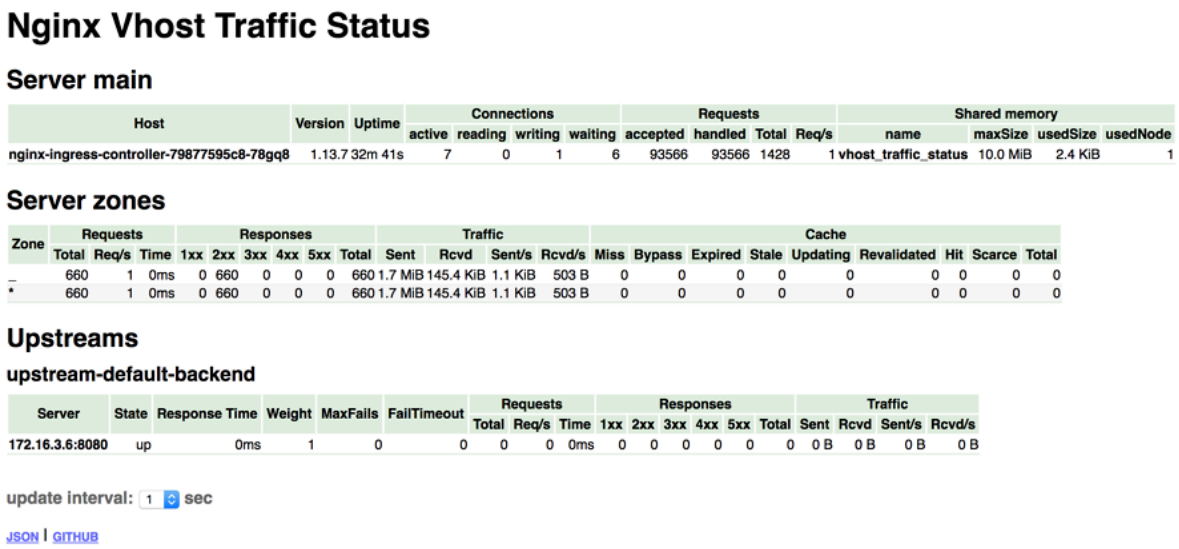
```
\"ingress-nginx\"},\"name\":\"nginx-configuration\",\"namespace\":\"
kube-system\"}}\n"
    }
 },
 "data": {
   "proxy-body-size": "20m",
   "enable-vts-status": "true"
 }
}
```

4. **Access the Ingress NGINX console from a local device.**

> **Note:**
> **By default, the VTS port is not exposed to ensure security. In this example, port forwarding is used for access.**

```
root@master # kubectl port-forward nginx-ingress-controller-
79877595c8-78gq8 -n kube-system 18080
Forwarding from 127.0.0.1:18080 -> 18080
Handling connection for 18080
```

5. **Use `http://localhost:18080/nginx_status` to access the VTS monitoring console.**



Nginx Vhost Traffic Status

Server main

Host	Version	Uptime	Connections				Requests				Shared memory			
			active	reading	writing	waiting	accepted	handled	Total	Req/s	name	maxSize	usedSize	usedNode
nginx-ingress-controller-79877595c8-78gq8	1.13.7	32m 41s	7	0	1	6	93566	93566	1428	1	vhost_traffic_status	10.0 MiB	2.4 KiB	1

Server zones

Zone	Requests			Responses					Traffic				Cache								
	Total	Req/s	Time	1xx	2xx	3xx	4xx	5xx	Total	Sent	Rcvd	Sent/s	Rcvd/s	Miss	Bypass	Expired	Stale	Updating	Revalidated	Hit	Scarce	Total
_	660	1	0ms	0	660	0	0	0	660	1.7 MiB	145.4 KiB	1.1 KiB	503 B	0	0	0	0	0	0	0	0	0
*	660	1	0ms	0	660	0	0	0	660	1.7 MiB	145.4 KiB	1.1 KiB	503 B	0	0	0	0	0	0	0	0	0

Upstreams

upstream-default-backend

Server	State	Response Time	Weight	MaxFails	FailTimeout	Requests			Responses						Traffic			
						Total	Req/s	Time	1xx	2xx	3xx	4xx	5xx	Total	Sent	Rcvd	Sent/s	Rcvd/s
172.16.3.6:8080	up	0ms	1	0	0	0	0	0ms	0	0	0	0	0	0	0 B	0 B	0 B	0 B

update interval: 1 sec

JSON | GITHUB

3.4.6.4 Ingress support

You can define Ingress rules for Kubernetes clusters in Container Service to implement control over load balancing.

In Kubernetes clusters, an Ingress is a collection of rules that allow inbound connections to reach the cluster services and provides you with Layer-7 load balancing capabilities. An Ingress can be configured to provide services with externally-reachable URLs, load balance traffic, terminate SSL, and offer name-based virtual hosting.

Prerequisites

In this example, an NGINX application is created to test a complex Ingress. You must create an NGINX deployment and multiple services to check the Ingress functionality. When performing the test, replace the services in this example with your own services.

```
root@master # kubectl run nginx --image=registry.cn-hangzhou.aliyuncs.
com/acs/netdia:latest

    root@master # kubectl expose deploy nginx --name=http-svc --port=
80 --target-port=80
    root@master # kubectl expose deploy nginx --name=http-svc1 --port=
80 --target-port=80
    root@master # kubectl expose deploy nginx --name=http-svc2 --port=
80 --target-port80
    root@master # kubectl expose deploy nginx --name=http-svc3 --port=
80 --target-port=80
```

Simple Ingresses

You can run the following command to create a simple Ingress. All access requests to the */svc* **path are routed to the http-svc service.** `nginx.ingress.kubernetes.io/rewrite-target: /` **redirects the** */svc* **path to the** */* **path that can be recognized by back-end services.**

```
root@master # cat <<EOF | kubectl create -f -
apiVersion: extensions/v1beta1
kind: Ingress
metadata:
  name: simple
  annotations:
    nginx.ingress.kubernetes.io/rewrite-target: /
spec:
  rules:
  - http:
      paths:
      - path: /svc
        backend:
          serviceName: http-svc
          servicePort: 80
EOF
```

```
root@master # kubectl get ing
NAME            HOSTS         ADDRESS          PORTS     AGE
simple          *             101.37.192.211   80        11s
```

Go to `http://101.37.192.211/svc` **to access the NGINX service.**

Simple domain name-based fanout Ingresses

If you have multiple domain names providing different external services, you can use the following configurations to implement a simple domain name-based fanout Ingress:

```
root@master # cat <<EOF | kubectl create -f -
apiVersion: extensions/v1beta1
kind: Ingress
metadata:
  name: simple-fanout
spec:
  rules:
  - host: foo.bar.com
    http:
      paths:
      - path: /foo
        backend:
          serviceName: http-svc1
          servicePort: 80
      - path: /bar
        backend:
          serviceName: http-svc2
          servicePort: 80
  - host: foo.example.com
    http:
      paths:
      - path: /film
        backend:
          serviceName: http-svc3
          servicePort: 80
EOF
```

```
root@master # kubectl get ing
NAME            HOSTS         ADDRESS          PORTS     AGE
simple-fanout   *             101.37.192.211   80        11s
```

Then, you can use `http://foo.bar.com/foo` to access the `http-svc1` service, use `http://foo.bar.com/bar` to access the `http-svc2` service, and use `http://foo.example.com/film` to access the `http-svc3` service.

Note:

- **In the production environment, you must link the domain to the preceding returned IP address `101.37.192.211`.**
- **In the test environment, you can modify the `hosts` file to add a domain name mapping rule.**

```
101.37.192.211 foo.bar.com
```

```
    101.37.192.211 foo.example.com
```

Default domain name of a simple Ingress

You can access Container Service even if you do not have the domain name of a simple Ingress. Container Service binds a default domain name to the Ingress service. You can use this default domain name to access the service. The domain name syntax is as follows: `*.[cluster-id].[region-id].alicontainer.com`. You can obtain the address from the cluster basic information page in the console.

You can use the following configuration to expose two services with the default domain name:

```
root@master # cat <<EOF | kubectl create -f -
apiVersion: extensions/v1beta1
kind: Ingress
metadata:
  name: shared-dns
spec:
  rules:
  - host: foo.[cluster-id].[region-id].alicontainer.com ##Replace with
 the default service access domain name of your cluster.
    http:
      paths:
      - path: /
        backend:
          serviceName: http-svc1
          servicePort: 80
  - host: bar.[cluster-id].[region-id].alicontainer.com ##Replace with
 the default service access domain name of your cluster.
    http:
      paths:
      - path: /
        backend:
          serviceName: http-svc2
          servicePort: 80
EOF
```

```
root@master # kubectl get ing
NAME            HOSTS         ADDRESS          PORTS     AGE
shared-dns   foo.[cluster-id].[region-id].alicontainer.com,bar.[
cluster-id].[region-id].alicontainer.com             47.95.160.171
80        40m
```

Then, you can use `http://foo.[cluster-id].[region-id].alicontainer.com/` to access the `http-svc1` service, and use `http://bar.[cluster-id].[region-id].alicontainer.com` to access the `http-svc2` service.

Configure a secure Ingress

Container Service allows you to manage multiple certificates to secure your services.

1. **Prepare your service certificate.**

   If no certificates are available, use the following method to generate a test certificate:

   Note:

   The domain name must be consistent with your Ingress configuration.

   ```
   root@master # openssl req -x509 -nodes -days 365 -newkey rsa:2048 -
   keyout tls.key -out tls.crt -subj "/CN=foo.bar.com/O=foo.bar.com"
   ```

   You can run the preceding command to generate a certificate file named *tls.crt* **and a private key file named** *tls.key***.**

   Then, you can use the certificate and private key to create a Kubernetes secret named foo.bar. You need to reference this secret when you create the Ingress.

   ```
   root@master # kubectl create secret tls foo.bar --key tls.key --cert
    tls.crt
   ```

2. **Create a secure Ingress.**

   ```
   root@master # cat <<EOF | kubectl create -f -
   apiVersion: extensions/v1beta1
   kind: Ingress
   metadata:
     name: tls-fanout
   spec:
     tls:
     - hosts:
       - foo.bar.com
       secretName: foo.bar
     rules:
     - host: foo.bar.com
       http:
         paths:
         - path: /foo
           backend:
             serviceName: http-svc1
             servicePort: 80
         - path: /bar
           backend:
             serviceName: http-svc2
             servicePort: 80
   EOF
   ```

   ```
   root@master # kubectl get ing
   NAME            HOSTS         ADDRESS          PORTS     AGE
   ```

```
tls-fanout      *               101.37.192.211   80        11s
```

3. **Follow the notes in Simple domain name-based fanout Ingresses to configure the `hosts` file or set the domain name to access the TLS service.**

   You can use `http://foo.bar.com/foo` to access the `http-svc1` service, and use `http://foo.bar.com/bar` to access the `http-svc2` service.

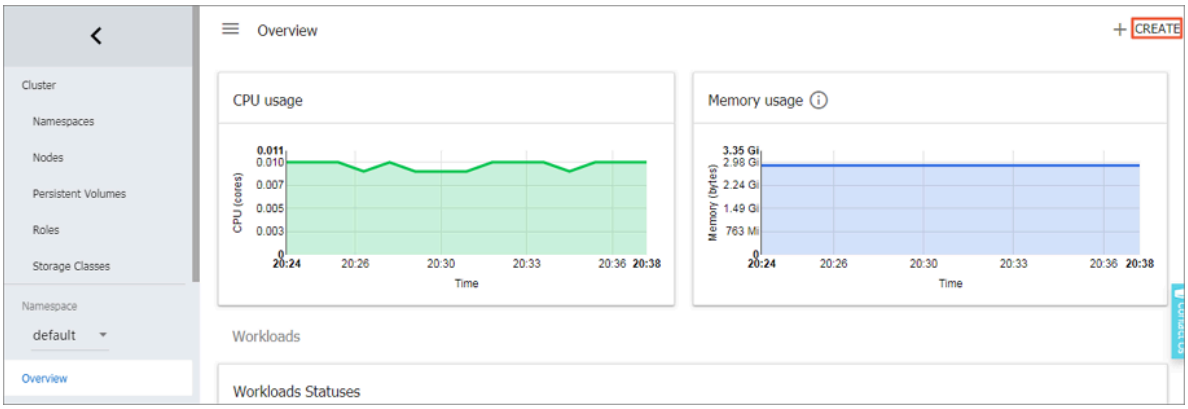
   You can also access the HTTPS service over HTTP. An Ingress will automatically redirect HTTP access requests to configured HTTPS addresses by default. For example, you will be automatically redirected to `https://foo.bar.com/foo` after accessing `http://foo.bar.com/foo`.

Deploy an Ingress from the Kubernetes dashboard

1. **Save the following YML code to the `nginx-ingress.yml` file.**

```
apiVersion: extensions/v1beta1
kind: Ingress
metadata:
  name: simple
spec:
  rules:
  - http:
      paths:
      - path: /svc
        backend:
          serviceName: http-svc
          servicePort: 80
```

2. *Log on to the Container Service console*.
3. **On the Clusters page, click Console in the Actions column corresponding to a cluster. The Kubernetes dashboard page appears.**
4. **Click Create in the upper-right corner to create an application.**



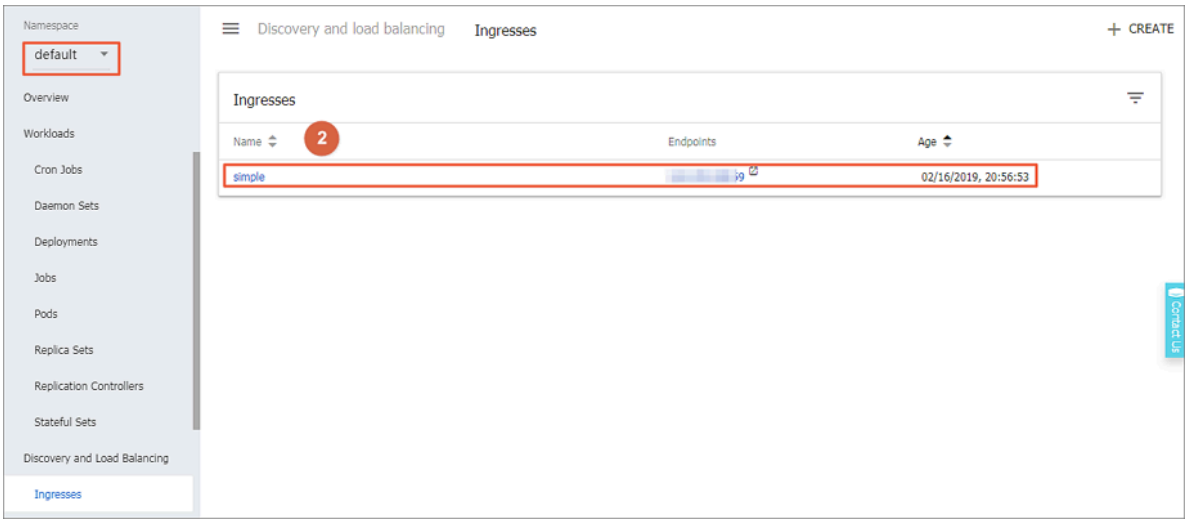
5. **Click the Create from File tab. Select the `nginx-ingress.yml` file you saved.**

6. **Click Upload.**

   An Ingress is then created on the `http-svc` service.

7. **Locate the default namespace in the Kubernetes dashboard and select Access Permission.**

   You can view the Ingress resource you just created and its access address.



8. **Enter the address in your browser to access the created `http-svc` service.**

3.4.6.5 Ingress configurations

Container Service provides Ingress controller components. Integrated with Apsara Server Load Balancer, these components provide Kubernetes clusters with flexible and reliable Ingress service.

An Ingress orchestration template is provided below. When you configure an Ingress through the console, you need to configure annotations and may need to create dependencies. For more information, see *Create Ingresses through the console*, *Ingress support*, **and** *Kubernetes Ingress*. **You can also create ConfigMaps to configure Ingresses. For more information, see** *https://kubernetes.github.io/ingress-nginx/user-guide/nginx-configuration/configmap/*.

```
apiVersion: extensions/v1beta1
kind: Ingress
metadata:
  annotations:
    nginx.ingress.kubernetes.io/service-match: 'new-nginx: header("foo
", /^bar$/)' #Canary release rule. In this example, the request header
 is used.
    Nginx. ingress. kubernetes. IO/service-weight: 'New-nginx: 50, old
-nginx: 50' #The route weight.
  creationTimestamp: null
  generation: 1
  name: nginx-ingress
```

```
  selfLink: /apis/extensions/v1beta1/namespaces/default/ingresses/
nginx-ingress
spec:
  rules: ##The Ingress rule.
  - host: foo.bar.com
    http:
      paths:
      - backend:
          serviceName: new-nginx
          servicePort: 80
        path: /
      - backend:
          serviceName: old-nginx
          servicePort: 80
        path: /
tls:            ## Enable TLS for secure routing.
  - hosts:
    - *.xxxxxx.cn-hangzhou.alicontainer.com
    - foo.bar.com
    secretName: nginx-ingress-secret     ##The Secret name.
status:
  loadBalancer: {}
```

Annotations

For each Ingress, you can configure its annotations, Ingress controller, and rules, such as the route weight, canary release rule, and rewrite rules. For more information about annotations, see *https://kubernetes.github.io/ingress-nginx/user-guide/nginx-configuration/annotations/*.

For example, the following rewrite annotation, `nginx.ingress.kubernetes.io/rewrite-target: /`**, indicates that** */path* **is redirected to the root path** */***, which can be recognized by the backend service.**

Rules

Ingress rules are used to manage external access to the services in the cluster and can be HTTP or HTTPS rules. You can configure the following items in rules: domain name (virtual hostname), URL path, service name, and port.

For each rule, you need to set the following parameters:

- **Domain: The test domain or virtual hostname of your service, such as** `foo.bar.com`**.**
- **Path: The URL path of your service. Each path is associated with a backend service. Server Load Balancer only forwards traffic to the backend if the incoming request matches the domain and path.**

- **Service: Specify the service in the form of `service:port`. You also need to specify a route weight for each service. The Ingress routes traffic to the matching service based on the route weight.**
  - **Name: The name of the backend service.**
  - **Port: The port of the service.**
  - **Weight: The route weight of the service in the service group.**

> **Note:**
>
> 1. **The weight is a percentage value. For example, you can set two services to the same weight of 50%.**
> 2. **A service group includes services that have the same domain and path defined in the Ingress configuration. If no weight is set for a service, the default value, 100, is used.**

Canary release

Container Service supports multiple traffic splitting approaches to suit scenarios such as canary release and A/B testing.

> **Note:**
> **Currently, only Ingress controllers of `0.12.0-5` and later versions support traffic splitting.**

1. **Traffic splitting based on request header**
2. **Traffic splitting based on cookie**
3. **Traffic splitting based on query parameter**

After canary release is configured, only requests that match certain rules are routed to the corresponding service. If the weight of the corresponding service is lower than 100%, requests that match certain rules are routed to one of the services in the service group based on the weight.

TLS

You can use a Secret that contains a TLS private key and certificate to encrypt the Ingress. This ensures secure routing. The TLS Secret must contain a certificate named tls.crt and a private key named tls.key. For more information about how TLS works, see *TLS*. For how to create a Secret, see *Configure a secure Ingress*.

Labels

You can add labels to the Ingress.

3.4.6.6 Create Ingresses through the console

Integrated with the Ingress service, the Container Service console allows you to create Ingresses to manage external access to the services in your cluster flexibly.

Prerequisites

- **You have created a Kubernetes cluster and an Ingress controller is running normally in the cluster. For more information about creating clusters, see** *Create a Kubernetes cluster***.**
- **You can use kubectl to connect to the master node. For more information, see** *Connect to a Kubernetes cluster through kubectl***.**
- **The image address used in this example requires Internet access. You can use the image address of your own cluster to replace it. Or you can build and push the image used here to the repository and pull it from the repository when you use it.**

Step 1: Create a Deployment and a Service

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Applications > Deployments to go to the Deployments page.**
3. **Click Create from Template in the upper-right corner.**
4. **Select the cluster and namespace. Select a sample template or enter a custom template, and then click Create.**

   This example creates two Nginx applications. One is named old-nginx and the other new-nginx.

   The orchestration template for old-nginx is as follows:

```
apiVersion: extensions/v1beta1
kind: Deployment
metadata:
  name: old-nginx
spec:
  replicas: 2
  selector:
    matchLabels:
      run: old-nginx
  template:
    metadata:
      labels:
        run: old-nginx
    spec:
```

```
      containers:
      - image: registry.cn-hangzhou.aliyuncs.com/xianlu/old-nginx
        imagePullPolicy: Always
        name: old-nginx
        ports:
        - containerPort: 80
          protocol: TCP
      restartPolicy: Always
---
apiVersion: v1
kind: Service
metadata:
  name: old-nginx
spec:
  ports:
  - port: 80
    protocol: TCP
    targetPort: 80
  selector:
    run: old-nginx
  sessionAffinity: None
  type: NodePort
```

The orchestration template for new-nginx is as follows:

```
apiVersion: extensions/v1beta1
kind: Deployment
metadata:
  name: new-nginx
spec:
  replicas: 1
  selector:
    matchLabels:
      run: new-nginx
  template:
    metadata:
      labels:
        run: new-nginx
    spec:
      containers:
      - image: registry.cn-hangzhou.aliyuncs.com/xianlu/new-nginx
        imagePullPolicy: Always
        name: new-nginx
        ports:
        - containerPort: 80
          protocol: TCP
      restartPolicy: Always
---
apiVersion: v1
kind: Service
metadata:
  name: new-nginx
spec:
  ports:
  - port: 80
    protocol: TCP
    targetPort: 80
  selector:
    run: new-nginx
  sessionAffinity: None
```

```
  type: NodePort
```

5. **In the left-side navigation pane, choose Ingresses and Load Balancing > Services to go to the Services page.**

   After the services are created, you can see them on the Services page.

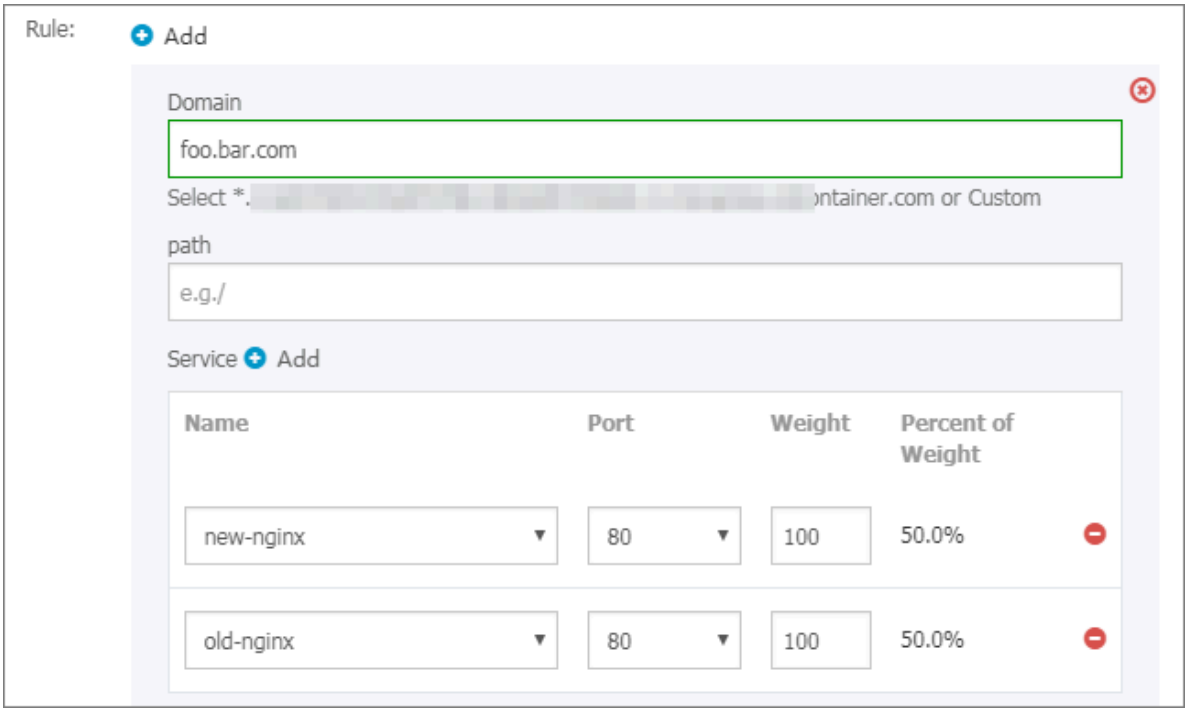
Step 2. Create an Ingress

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Ingresses and Load Balancing > Ingresses to go to the Ingresses page.**
3. **Select the cluster and namespace, and then click Create in the upper-right corner.**
4. **In the dialog box that appears, enter the Ingress name. In this example, nginx-ingress.**
5. **Configure rules.**

   Ingress rules are used to manage external access to the services in the cluster and can be HTTP or HTTPS rules. You can configure the following items in rules:

domain name (virtual hostname), URL path, service name, port, and weight. For more information, see *Ingress configurations*.

This example adds a complex rule and configures the default test domain and virtual hostname for the cluster. Traffic routing is based on domains.



Simple fanout based on domains

This example uses a virtual hostname as the test domain to provide services to the public. Route weights are set for both services and canary release is

configured for one of the services. In the production environment, you can use the domain that has obtained an ICP filing to provide services.

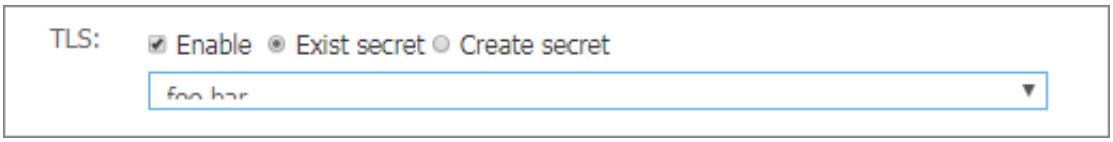
- **Domain: In this example, the test domain `foo.bar.com` is used.**

  You need to modify the hosts file and add a domain mapping rule.

  ```
  118.178.108.143 foo.bar.com # The IP address of the Ingress
  ```

- **Service: Configure the following parameters: path, name, port, and weight.**
  - **Path: Specify the URL path of the service. In this example, the root path / is used.**
  - **Name: In this example, specify both services, nginx-new and nginx-old.**
  - **Port: In this example, open port 80.**
  - **Weight: Set the weight of each service. The weight is a percentage value. The default value is 100. In this example, the old and new service have the same weight, 50%.**

6. **Configure TLS. Select Enable and configure secure routing. For more information, see** *Configure a secure Ingress*.
   - **You can use an existing Secret.**



     a. **Log on to the master node and create** *tls.key* **and** *tls.crt*.

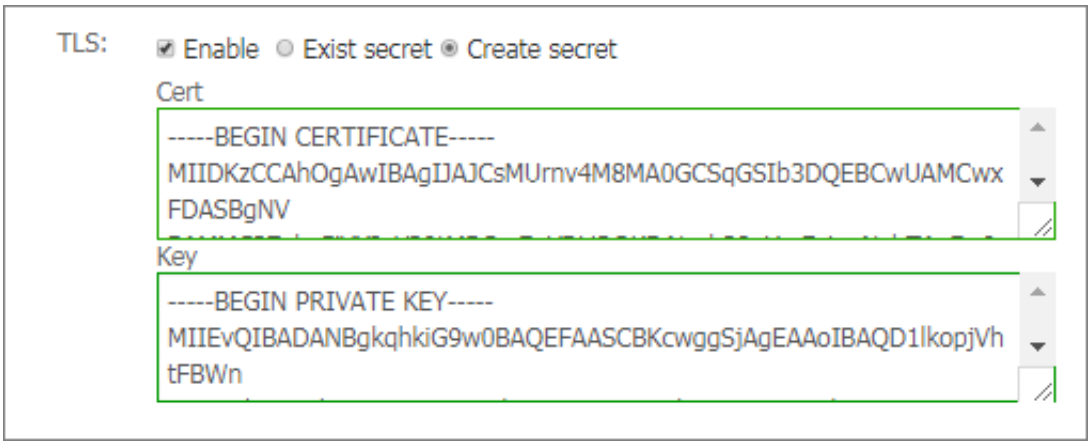
```
openssl req -x509 -nodes -days 365 -newkey rsa:2048 -keyout tls
.key -out tls.crt -subj "/CN=foo.bar.com/O=foo.bar.com"
```

     b. **Create a Secret.**

```
kubectl create secret tls foo.bar --key tls.key --cert tls.crt
```

     c. **Run the** `kubectl get secret` **command and check whether the Secret is created. You can select the newly created** *foo.bar* **Secret.**

   - **You can also use the TLS private key and certificate to create a new Secret.**



     a. **Log on to the master node and create** *tls.key* **and** *tls.crt*.

```
openssl req -x509 -nodes -days 365 -newkey rsa:2048 -keyout tls
.key -out tls.crt -subj "/CN=foo.bar.com/O=foo.bar.com"
```

     b. **Run the** `vim tls.key` **and** `vim tls.crt` **commands to get the private key and certificate.**

     c. **Copy the certificate and private key to the Cert and Key fields.**

7. **Configure canary release.**

Note:

Currently, only Ingress controllers of `0.12.0-5` and later versions support traffic splitting.

Container Service supports multiple traffic splitting approaches to suit scenarios such as canary release and A/B testing.

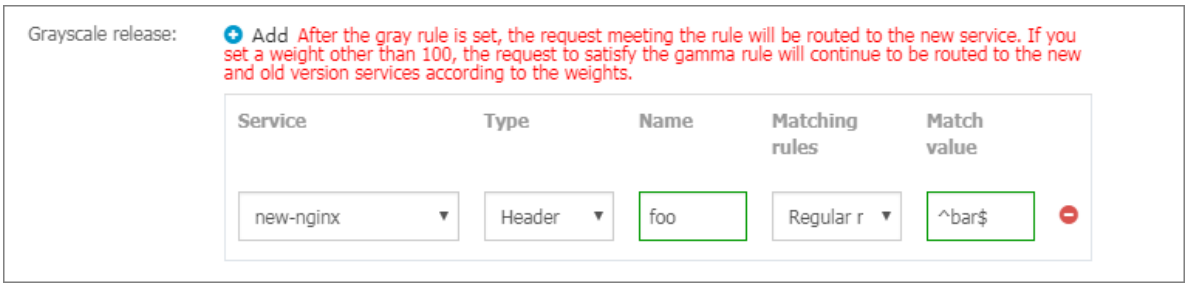
a. Traffic splitting based on request header

b. Traffic splitting based on cookie

c. Traffic splitting based on query parameter

After canary release is configured, only requests that match certain rules are routed to the new service new-nginx. If the weight of new-nginx is lower than 100%, requests that match certain rules are routed to this service based on the weight.

This example sets the rule on the request header to `foo=^bar$`. Only requests that contain this header can access new-nginx.



- **Service: The service to be accessed.**
- **Type: The type of the matching rule, such as Header, Cookie, or Query.**
- **Name and Matching Value: Custom field. The name and matching value are a key-value pair.**
- **Matching Rules: Regular expressions and exact matches are supported.**

8. **Configure annotations.**

   Click Rewrite Annotation to add a redirection annotation for the Ingress. `nginx.ingress.kubernetes.io/rewrite-target: /` indicates that *`/path`* is redirected to the root path `/`, which can be recognized by the backend service.

   Note:
   In this example, no access path is configured for the service. Therefore, you do not need to configure rewrite annotations. Rewrite annotations enable Ingress

to forward path as root path to the backend. This helps avoid 404 errors that are caused by incorrect paths.

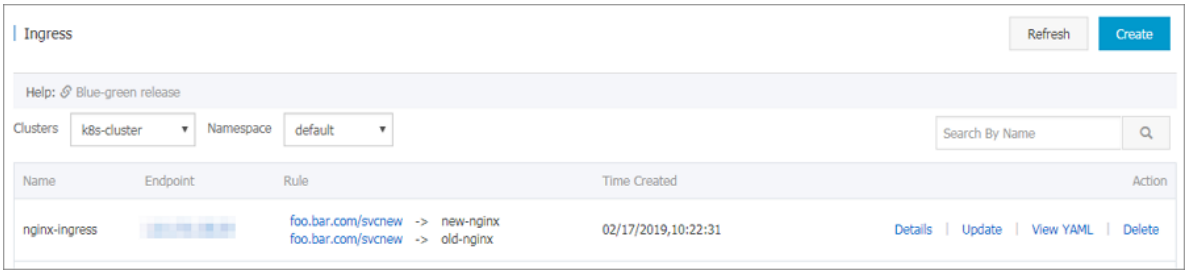
You can also click Add to enter annotation names and values in key-value pairs. For more information about Ingress annotations, see *https://kubernetes.github.io/ingress-nginx/user-guide/nginx-configuration/annotations/*.

9. Add labels.

   Labels are used to indicate the features of an Ingress.

10. Click Create.

    You can find the nginx-ingress Ingress on the Ingresses page.



11. Click foo.bar.com and the Nginx welcome page appears.

    Click the route address that points to new-nginx. You will find that you are directed to old-nginx.

> **Note:**
> By default, when you enter the route address in the browser, requests with headers that do not contain `foo=^bar$` are directed to old-nginx.

12. Log on to the master node by using SSH. Run the following commands to simulate requests with specific headers and check the results.

```
curl -H "Host: foo.bar.com" http://47.107.20.35
old
 curl -H "Host: foo.bar.com" http://47.107.20.35
old
 curl -H "Host: foo.bar.com" http://47.107.20.35 # Similar to
browser requests
old
 curl -H "Host: foo.bar.com" -H "foo: bar" http://47.107.20.35 #
Simulate a request with a specific header. The results are returned
based on the route weight.
new
 curl -H "Host: foo.bar.com" -H "foo: bar" http://47.107.20.35
old
 curl -H "Host: foo.bar.com" -H "foo: bar" http://47.107.20.35
old
 curl -H "Host: foo.bar.com" -H "foo: bar" http://47.107.20.35
new
```

3.4.6.7 Update Ingresses

You can update Ingresses in the Container Service console.

Prerequisites

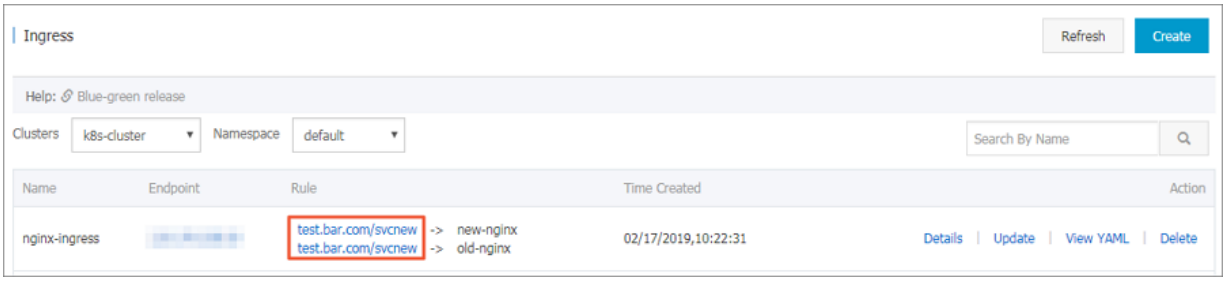
- You have created a Kubernetes cluster and an Ingress controller is running normally in the cluster. For more information about creating clusters, see *Create a Kubernetes cluster*.
- You have created an Ingress. For more information, see *Create Ingresses through the console*.

Procedure

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Ingresses and Load Balancing > Ingresses to go to the Ingresses page.
3. Select the cluster and namespace. Select the Ingress that you want to update and click Update in the Actions column.
4. In the dialog box that appears, modify the parameters and click OK. This example changes `foo.bar.com` to `test.bar.com`.

What's next

On the Ingresses page, you can find the changed Ingress rule.



3.4.6.8 Delete Ingresses

Prerequisites

- **You have created a Kubernetes cluster and an Ingress controller is running normally in the cluster. For more information about creating clusters, see** *Create a Kubernetes cluster*.
- **You have created an Ingress. For more information, see** *Create Ingresses through the console*.

Procedure

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Ingresses and Load Balancing > Ingresses.**
3. **Select the cluster and namespace. Select the Ingress that you want to delete and click Delete in the Actions column.**
4. **In the dialog box that appears, click OK.**

3.4.7 Config maps and secrets

3.4.7.1 Create ConfigMaps

In the Container Service console, you can create ConfigMaps on the ConfigMaps page or by using templates.

Create ConfigMaps on the ConfigMaps page

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Configuration > ConfigMaps to go to the ConfigMaps page.**
3. **Select the cluster and namespace, and then click Create.**

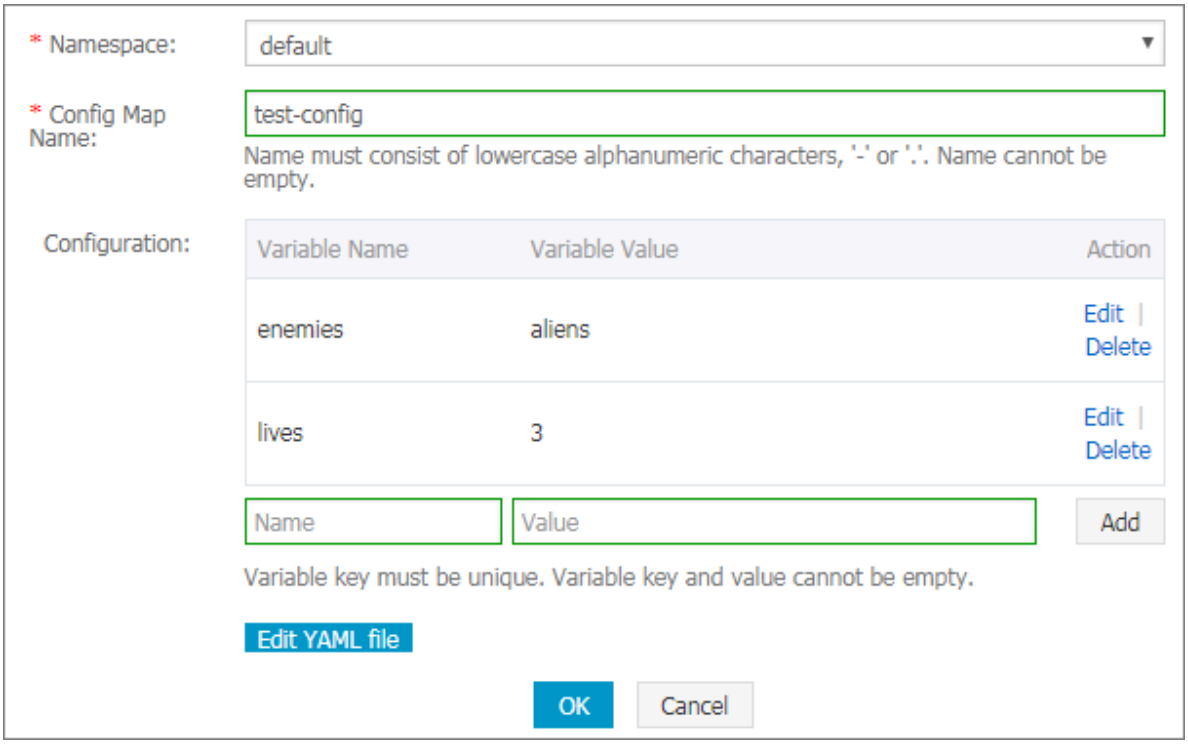
4. **Set the parameters and click OK.**

Table 3-9: Create a ConfigMap on the ConfigMaps page

Parameter	Description
Cluster	**The ID of the cluster that you have selected.**
Namespace	**The namespace that you have selected . ConfigMaps are a kind of Kubernetes resource object and must be divided into namespaces.**
ConfigMap Name	**Required. The name can contain lowercase letters, numbers, hyphens (-), and periods (.). Other resource objects need to reference ConfigMap names to obtain configuration information.**

Parameter	Description
ConfigMap	Enter the Name and Value, and then click Add to add this key-value pair. You can also click Edit ConfigMap, modify the parameters in the dialog box that appears, and then click OK.

This example creates two variables named enemies and lives, and sets their values to aliens and 3 respectively.



5. After the configuration is complete, click OK. You can find the test-config ConfigMap on the ConfigMaps page.

Create a ConfigMap from a template

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Applications > Deployments to go to the Deployments page.
3. Click Create from Template in the upper-right corner.

4. **On the Create from Template page, modify the sample template and click Create.**

   Table 3-10: Create a ConfigMap from a template

Parameter	Description
Cluster	**The cluster where the ConfigMap is created.**
Namespace	**The namespace that the ConfigMap belongs to. ConfigMaps are a kind of Kubernetes resource object and must be divided into namespaces.**
Sample Template	**You can choose `Custom` and write your own ConfigMap based on YAML syntax, or select the `Resource-ConfigMap` sample template. In the sample template, the ConfigMap is named aliyun-config and contains two variable files *`game.properties`* and *`ui.properties`*. You can make changes to the ConfigMap based on your needs.**

   After the deployment is complete, you can find the *`aliyun-config`* ConfigMap on the ConfigMaps page.

3.4.7.2 Use a ConfigMap in a Pod

You can use a ConfigMap in a Pod in the following scenarios:

- **Use a ConfigMap to define environment variables**
- **Use a ConfigMap to configure command line parameters**
- **Use a ConfigMap in volumes**

For more information, see *Configure a Pod to use a ConfigMap*.

Limits

To use a ConfigMap in a Pod, make sure that the ConfigMap and Pod are in the same cluster and namespace.

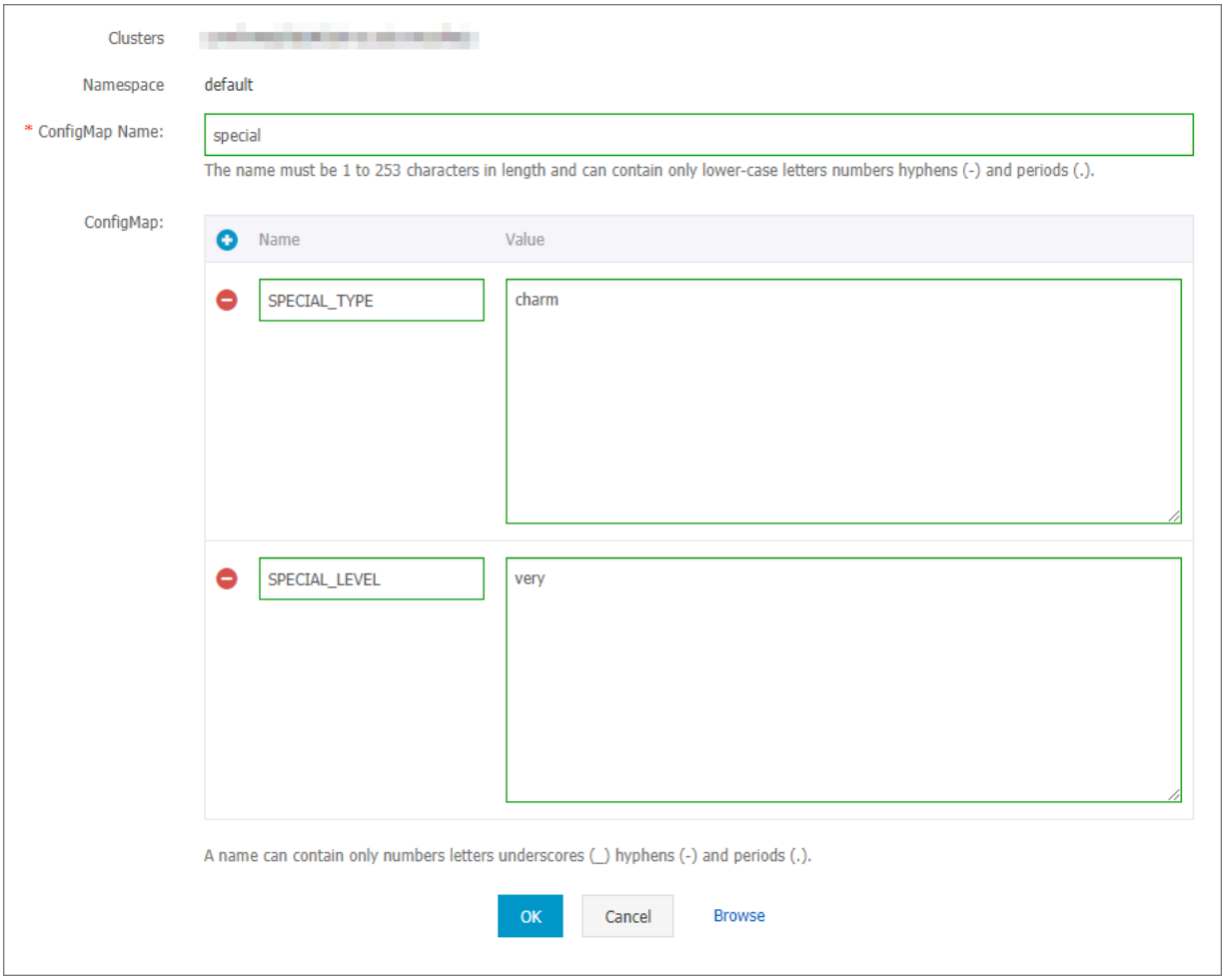
Create a ConfigMap

This example creates a ConfigMap named special_config, which consists of two key-value pairs: `SPECIAL_LEVEL: very` and `SPECIAL_TYPE: charm`.

You can use the following YAML template to create a ConfigMap.

```
apiVersion: v1
kind: ConfigMap
metadata:
  name: special-config
  namespace: default
data:
  SPECIAL_LEVEL: very
  SPECIAL_TYPE: charm
```

You can also log on to the Container Service console and choose Configuration > ConfigMaps in the left-side navigation pane. You can then click Create to create a ConfigMap.



Use ConfigMaps to define Pod environment variables

Define the value of a ConfigMap as an environment variable

You can log on to the Container Service console and choose Applications > Deployments in the left-side navigation pane. Click Create from Template, select and modify the Pod type template, and deploy the application. You can also go to the Kubernetes dashboard and choose Upload YAML or JSON File.

The following sample template creates a Pod and defines environment variables in the Pod. `valueFrom` is used to reference the value of SPECIAL_LEVEL to define an environment variable.

```
apiVersion: v1
kind: Pod
metadata:
  name: config-pod-1
spec:
  containers:
    - name: test-container
      image: busybox
      command: [ "/bin/sh", "-c", "env" ]
      env:
        - name: SPECIAL_LEVEL_KEY
          valueFrom:              ##Use valueFrom to denote that env
references the value of a ConfigMap.
            configMapKeyRef:
              name: special-config               ##The referenced
ConfigMap name.
              key: SPECIAL_LEVEL                 ##The referenced
ConfigMap key.
  restartPolicy: Never
```

To define the values of multiple ConfigMaps as environment variables, you only need to add multiple env parameters in the Pod definition.

Define the key-value pairs of a ConfigMap as environment variables

To define the key-value pairs of a ConfigMap as Pod environment variables, you can use the envFrom parameter. The keys in a ConfigMap are used as the names of the environment variables.

A sample template is provided as follows:

```
apiVersion: v1
kind: Pod
metadata:
  name: config-pod-2
spec:
  containers:
    - name: test-container
      image: busybox
      command: [ "/bin/sh", "-c", "env" ]
      envFrom:                ##Reference all the key-value pairs in
the special-config ConfigMap.
      - configMapRef:
        name: special-config
```

```
  restartPolicy: Never
```

Use a ConfigMap to configure command line parameters

You can use ConfigMaps to configure the commands or parameter values in a container by using the environment variable replacement syntax *$(VAR_NAME)*. A sample template is provided as follows:

```
apiVersion: v1
kind: Pod
metadata:
  name: config-pod-3
spec:
  containers:
    - name: test-container
      image: busybox
      command: [ "/bin/sh", "-c", "echo $(SPECIAL_LEVEL_KEY) $(
SPECIAL_TYPE_KEY)" ]
      env:
        - name: SPECIAL_LEVEL_KEY
          valueFrom:
            configMapKeyRef:
              name: special-config
              key: SPECIAL_LEVEL
        - name: SPECIAL_TYPE_KEY
          valueFrom:
            configMapKeyRef:
              name: special-config
              key: SPECIAL_TYPE
  restartPolicy: Never
```

Run the Pod and the output is as follows:

```
very charm
```

Use a ConfigMap in volumes

You can use a ConfigMap to define volumes. The following sample template specifies a ConfigMap name under volumes. This stores the key-value pair data to the mountPath path, which is /etc/config in this example. This generates configurat ion files that are named after the keys of the ConfigMap. The corresponding values of the ConfigMap are stored in these files.

```
apiVersion: v1
kind: Pod
metadata:
  name: config-pod-4
spec:
  containers:
    - name: test-container
      image: busybox
      command: [ "/bin/sh", "-c", "ls /etc/config/" ]   ##List the
names of files under this directory.
      volumeMounts:
        - name: config-volume
```

```
          mountPath: /etc/config
  volumes:
    - name: config-volume
      configMap:
        name: special-config
  restartPolicy: Never
```

Run the Pod and the keys of the ConfigMap are output:

```
SPECIAL_TYPE
SPECIAL_LEVEL
```

3.4.7.3 Update ConfigMaps

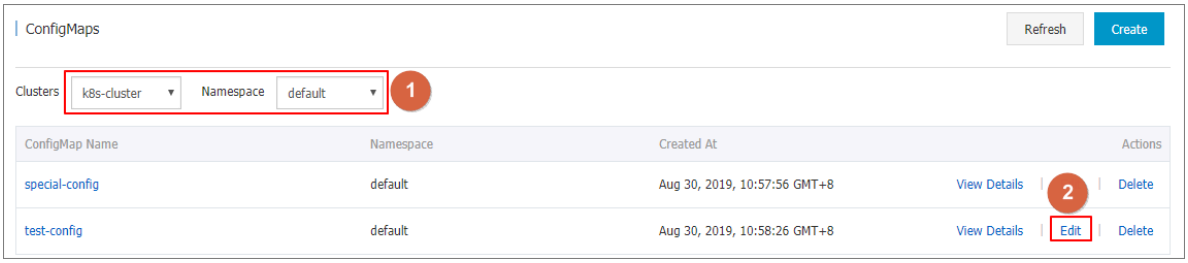
You can use multiple methods to update ConfigMaps.

Note

Updating a ConfigMap will affect applications that use this ConfigMap.

Update ConfigMaps on the ConfigMaps page

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Configuration > ConfigMaps to go to the ConfigMaps page.**
3. **Select the cluster and choose the ConfigMap that you want to update. Click Edit in the Actions column.**



4. **In the dialog box that appears, click OK.**
5. **Click Edit in the Actions column to edit the ConfigMap. Click Save to save your changes.**
6. **After the changes are saved, click OK.**

Update ConfigMaps through the Kubernetes dashboard

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Clusters > Clusters. Select the target cluster and click Dashboard in the Actions column.**

3. On the Kubernetes Dashboard page, choose Config and Storage > ConfigMaps in the left-side navigation pane. Select the target ConfigMap and click  > View/edit YAML.
4. Edit the ConfigMap in the dialog box that appears and then click Update.

3.4.7.4 Delete ConfigMaps

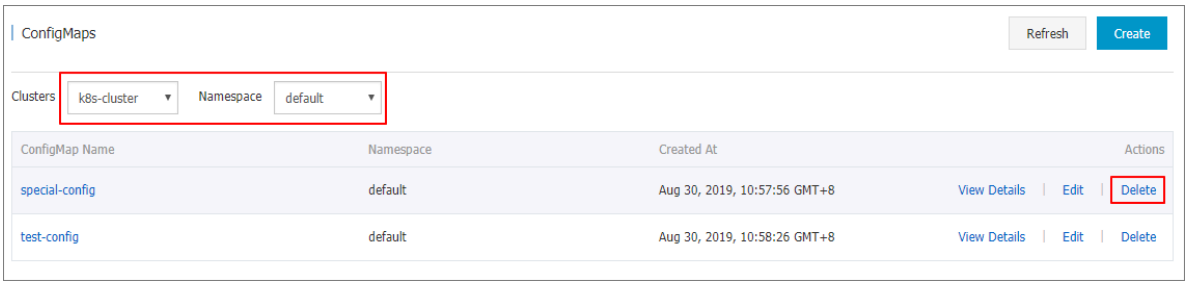
You can use multiple methods to delete ConfigMaps.

Note

Deleting a ConfigMap will affect applications that use this ConfigMap.

Delete ConfigMaps on the ConfigMaps page

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Configuration > ConfigMaps to go to the ConfigMaps page.
3. Select the cluster and choose the ConfigMap that you want to delete. Click Delete in the Actions column.



Delete ConfigMaps through the Kubernetes dashboard

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Clusters > Clusters. Select the target cluster and click Dashboard in the Actions column.
3. On the Kubernetes Dashboard page, choose Config and Storage > ConfigMaps in the left-side navigation pane. Select the target ConfigMap and click Actions > Delete.
4. In the dialog box that appears, click Delete.

3.4.7.5 Create Secrets

You can create Secrets for applications in the Container Service console.

Prerequisites

You have created a Kubernetes cluster.

Context

We recommend that you use Secrets to store sensitive information in Kubernetes clusters, such as passwords and certificates.

Secrets have multiple types as follows:

- **Service Account: This type of Secret is automatically created by Kubernetes and used to access the Kubernetes API. It is automatically mounted to the Pod directory** `/run/secrets/kubernetes.io/serviceaccount`**.**
- **Opaque: This type of Secret is encoded in base64 format and used to store sensitive information such as passwords and certificates.**

By default, you can only create Opaque Secrets in the Container Service console. Opaque data is map type data, which requires the value to be encoded in base64 format. You can encode plain text data into base64 format through the Container Service console.

You can also manually create Secrets by using command lines. For more information, see *Kubernetes secrets***.**

Procedure

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Configuration > Secrets to go to the Secrets page.**
3. **Select the cluster and namespace. Click Create in the upper-right corner.**
4. **Configure the Secret and click OK.**

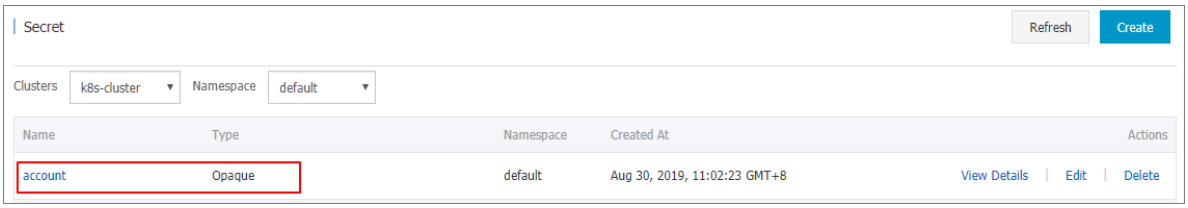
Note:

If you want to enter Secret data in plain text, select the Encode Data Values Using Base64 check box.

Table 3-11: Secret parameters

Parameter	Description
Name	The Secret name. The name must be 1 to 253 characters in length and can only contain lowercase letters, numbers, hyphens (-), or periods (.).
Data	The Secret data. Click Add and enter the Secret name and value, namely, the key-value pair in the dialog box that appears. In this example, the Secret contains two values: `username: admin` and`passwrod: 1f2d1e2e67df`.

5. You can view the newly created Secret in the Secrets list.



3.4.7.6 Edit Secrets

You can directly edit Secrets in the Container Service console.

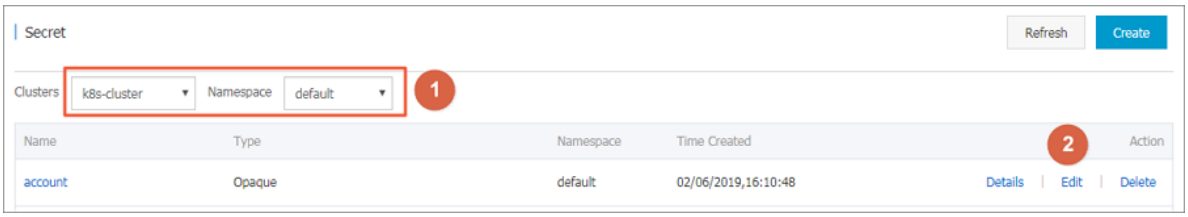
Prerequisites

- You have created a Kubernetes cluster.
- You have created a Secret. For more information, see *Create Secrets*.

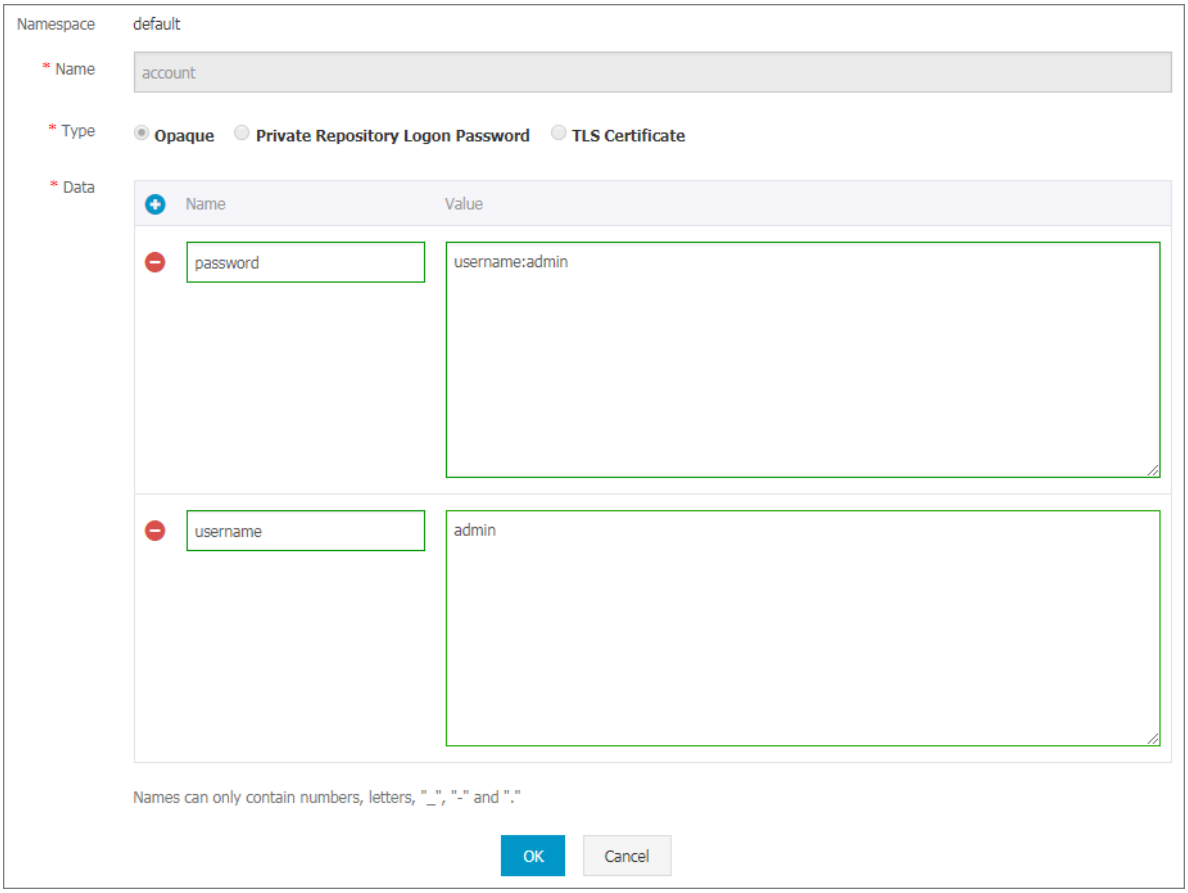
Procedure

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Configuration > Secrets to go to the Secrets page.

3. Select the cluster and namespace. Choose the Secret that you want to edit and Click Edit in the Actions column.



4. On the Edit Secret page, edit the Secret based on your needs.



5. Click OK.

3.4.7.7 Delete Secrets

You can delete Secrets in the Container Service console.

Prerequisites

- You have created a Kubernetes cluster.
- You have created a Secret. For more information, see *Create Secrets*.

Context

Note:

Do not delete Secrets that were generated during the cluster creation process.

Procedure

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Configuration > Secrets to go to the Secrets page.**
3. **Select the cluster and namespace. Choose the Secret that you want to delete and Click Delete in the Actions column.**
4. **In the dialog box that appears, click OK to delete the Secret.**

3.4.8 Templates

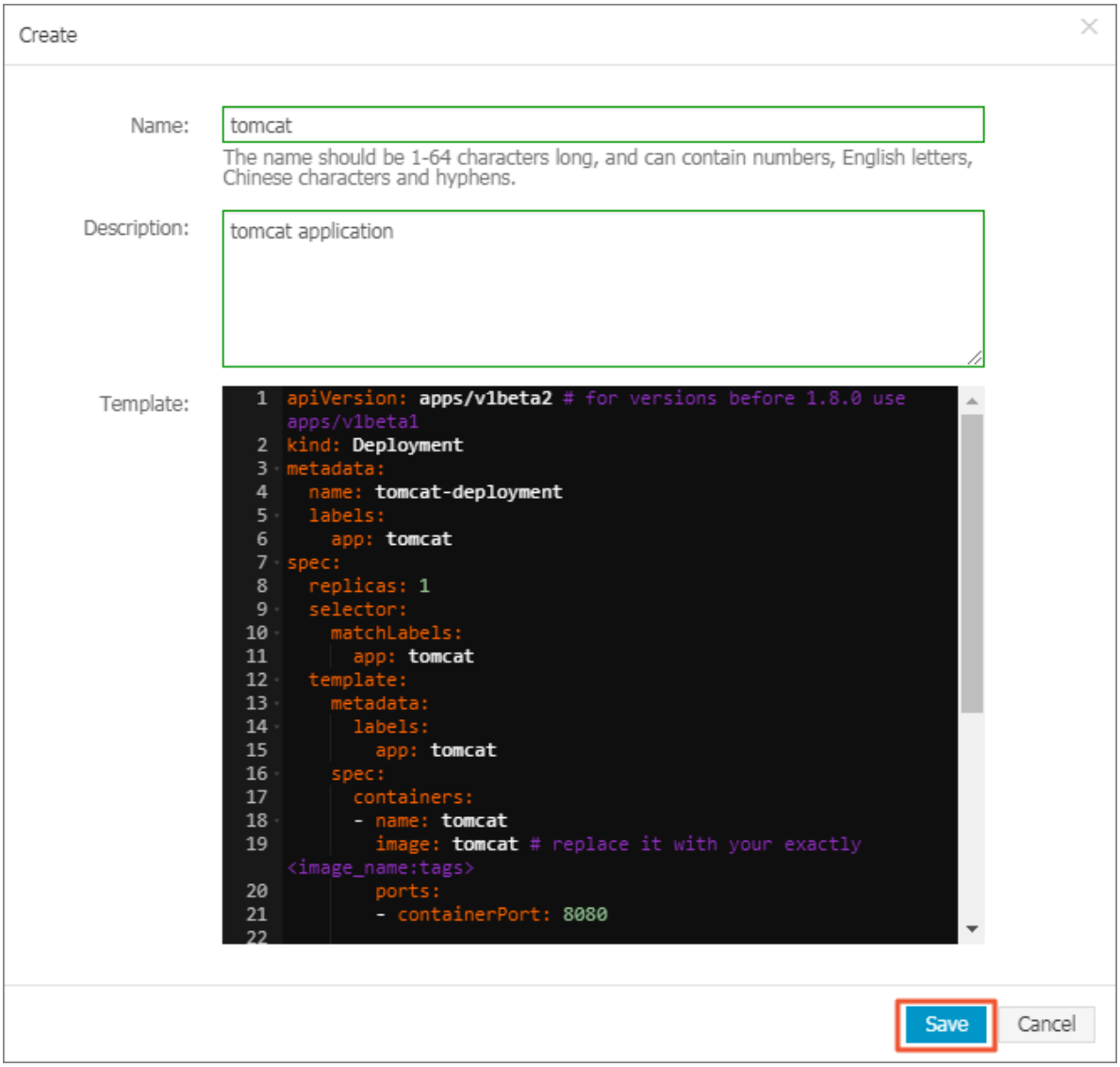
3.4.8.1 Create orchestration templates

You can use multiple methods to create orchestration templates through the Container Service console.

Procedure

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Marketplace > Orchestration Templates and click Create in the upper-right corner.**

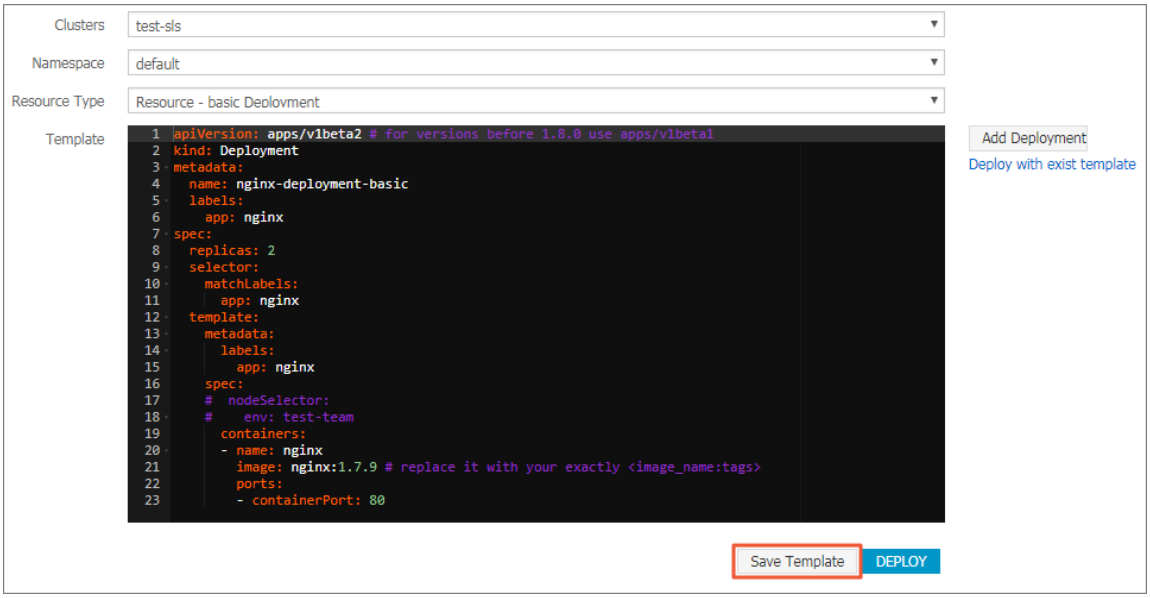
3. In the dialog box that appears, configure the orchestration template, and then click Save. This example demonstrates how to create a template of a Tomcat application that contains a Deployment and a Service.
   - Name: The template name.
   - Description: The description of the template. This parameter is optional.
   - Template: Configure the template based on YAML syntax. The template can contain multiple resource objects that are separated by `---`.



4. After the template is created, you are redirected to the Templates page. You can find the template on the My Templates tab.

5. Optional: You can also choose Applications > Deployments in the left-side navigation pane, and click Create from Template to go to the Create from Template page. You can modify a built-in orchestration template provided by Container Service and save it as your custom template.

   a) Select a built-in template and click Save Template.



   b) In the dialog box that appears, specify the name, description, and content. Click Save to save the template.

> **Note:**
> You can modify the built-in template based on your needs.

   c) In the left-side navigation pane, choose Marketplace > Orchestration Templates. You can find the newly created template on the My Templates tab.

What's next

You can use the orchestration templates on the My Templates tab to create applications quickly.

3.4.8.2 Edit an orchestration template

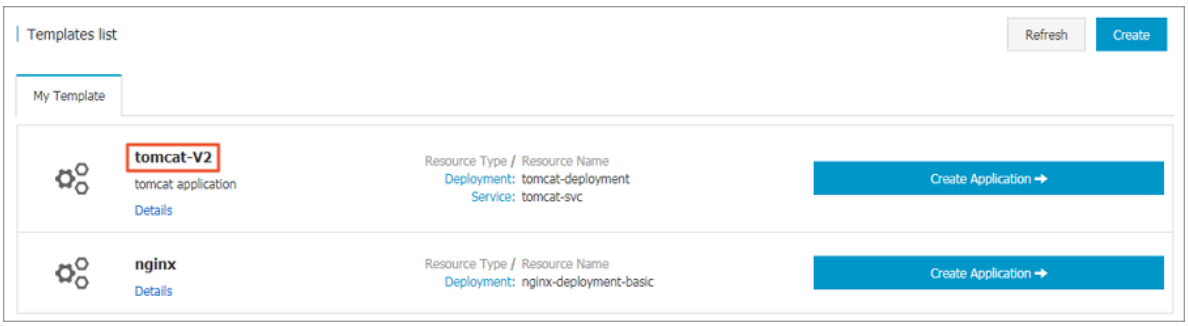
You can edit existing orchestration templates.

Prerequisites

You have created an orchestration template. For more information, see *Create orchestration templates*.

Procedure

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Market > Orchestration Templates. The Templates page appears. You can view existing orchestration templates on the My Templates tab page.**
3. **Select a template and click Details.**
4. **On the template details page, click Edit in the upper-right corner.**
5. **In the Edit Template dialog box that appears, edit the name, description, and template, and click Save.**
6. **Return to the Templates page. You can view the template that you modified on the My Templates tab page.**



3.4.8.3 Save an existing orchestration template as a new one

You can save existing templates as new ones.

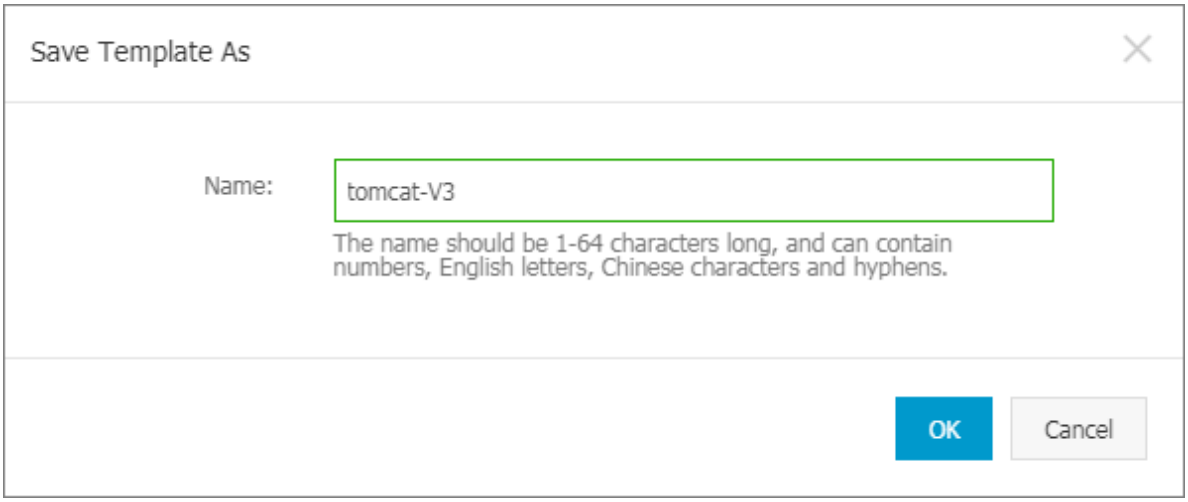
Prerequisites

You have created an orchestration template. For more information, see *Create orchestration templates*.

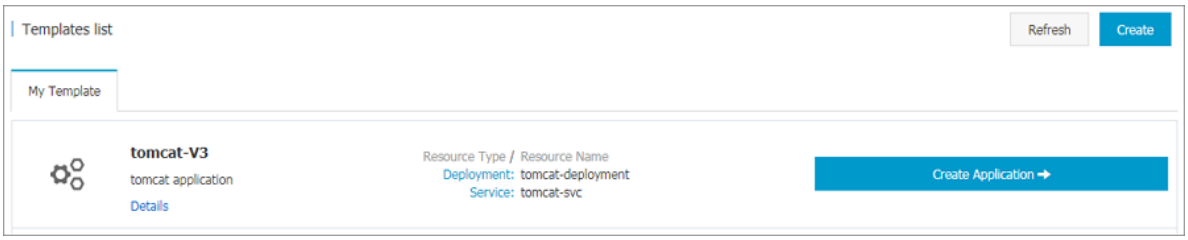
Procedure

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Market > Orchestration Templates. The Templates page appears. You can view existing orchestration templates on the My Templates tab page.**

3. Select a template and click Details.
4. On the template details page, modify the template and click Save As in the upper-right corner.
5. In the dialog box that appears, enter the template name and click OK.



6. Return to the Templates page. The saved template is displayed on the My Templates tab page.



3.4.8.4 Download an orchestration template

You can download existing orchestration templates.

Prerequisites

You have created an orchestration template. For more information, see *Create orchestration templates*.

Procedure

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Market > Orchestration Templates. The Templates page appears. You can view existing orchestration templates on the My Templates tab page.
3. Select a template and click Details.

4. Click Download in the upper-right corner of the template details page. A template file with the yml suffix is downloaded immediately.

3.4.8.5 Delete an orchestration template

Prerequisites

You have created an orchestration template. For more information, see *Create orchestration templates*.

Procedure

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Market > Orchestration Templates. The Templates page appears. You can view existing orchestration templates on the My Templates tab page.
3. Select a template and click Details.
4. Click Delete in the upper-right corner of the template details page.
5. In the message that appears, click OK.

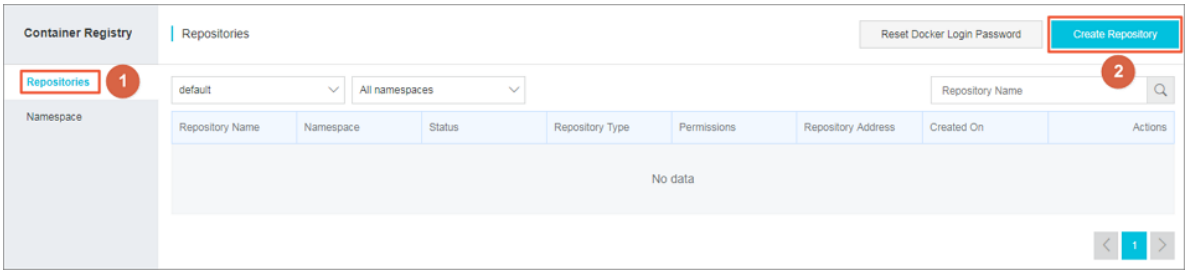
3.4.9 Images

3.4.9.1 Create a repository

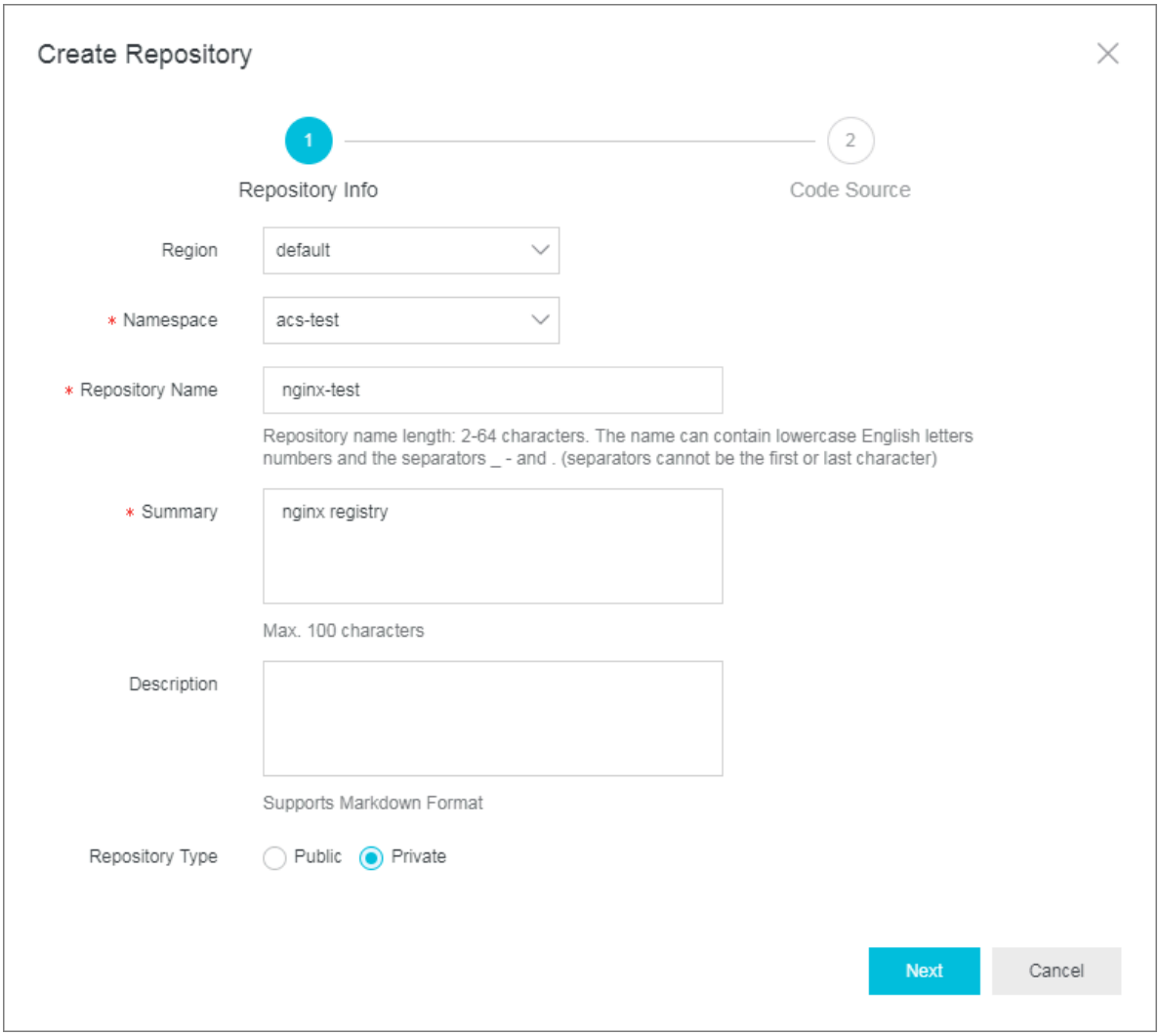
You can create a repository in the Container Registry console and upload your container images to the created repository.

Procedure

1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Market > Images.
3. In the left-side navigation pane of the Container Registry console, click Repositories. Then, click Create Repository in the upper-right corner of the Repositories page.

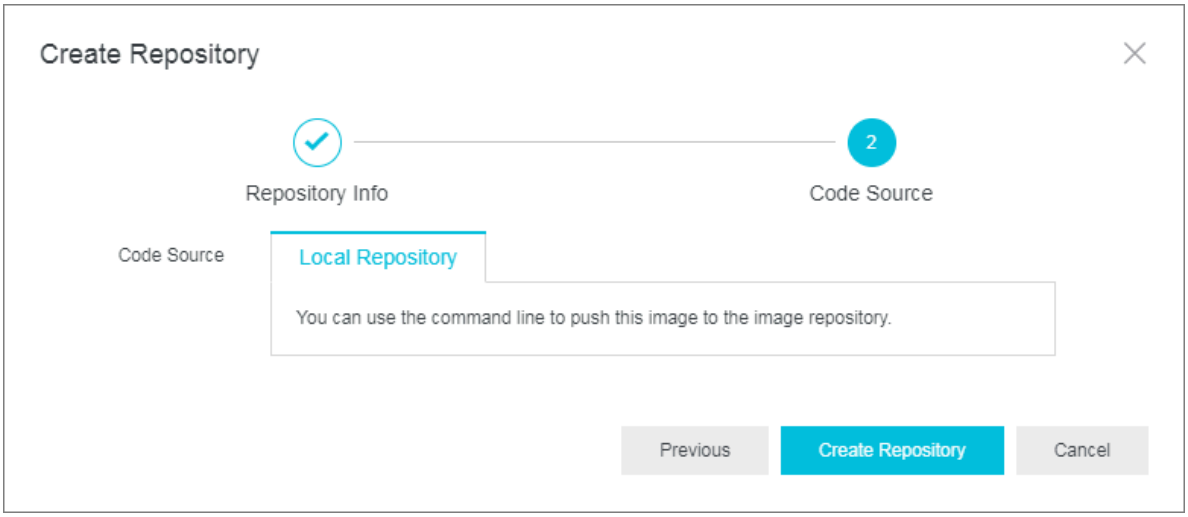


4. In the Create Repository dialog box that appears, set the repository parameters and click Next.
   - Region: Retain the default value. The repository must be deployed in the same region as the cluster.
   - Namespace: Retain the default value. The namespace must be the same as that of the department in the Container Service console.
   - Repository Name: Enter the repository name. The name must be 2 to 64 characters in length, and can contain lowercase letters, digits, underscores (_), hyphens (-), and periods(.). The name must not start with or end with a separator.
   - Summary: Enter summary information about the repository.
   - Description: Optional. Enter a description. The description can contain up to 100 characters.
   - Repository Type: Set this parameter to Public or Private.

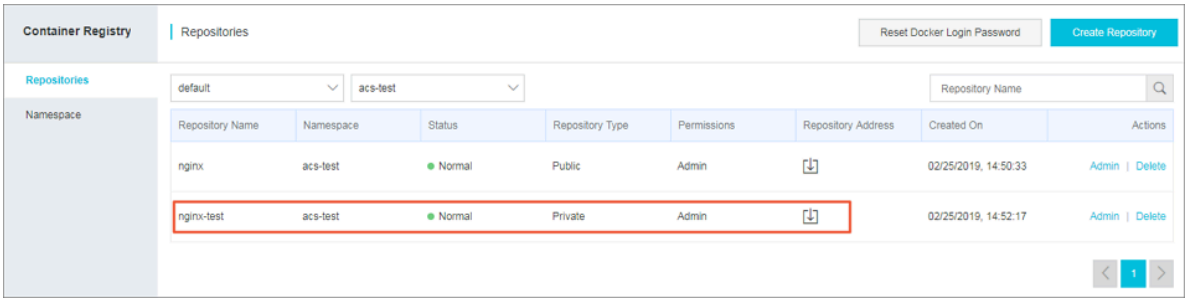


5. Set the code source, and then click Create Repository.

> **Note:**
> Only local repositories are available. You can run commands to upload images to the repository.

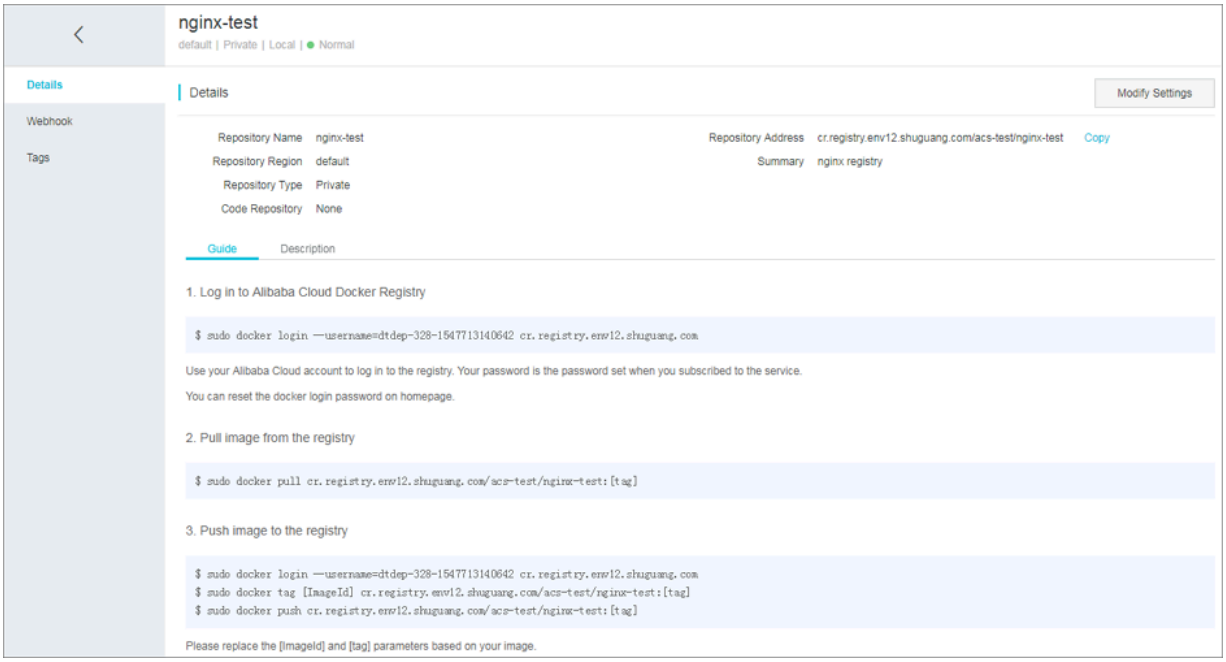


6. The repository is displayed in the repository list.



What's next

After the repository is created, click Admin in the Actions column corresponding to the repository. On the Details page that appears, view how to use the repository.



3.4.9.2 Create a namespace

You can create a namespace as a collection of repositories in the Container Registry console.

Context

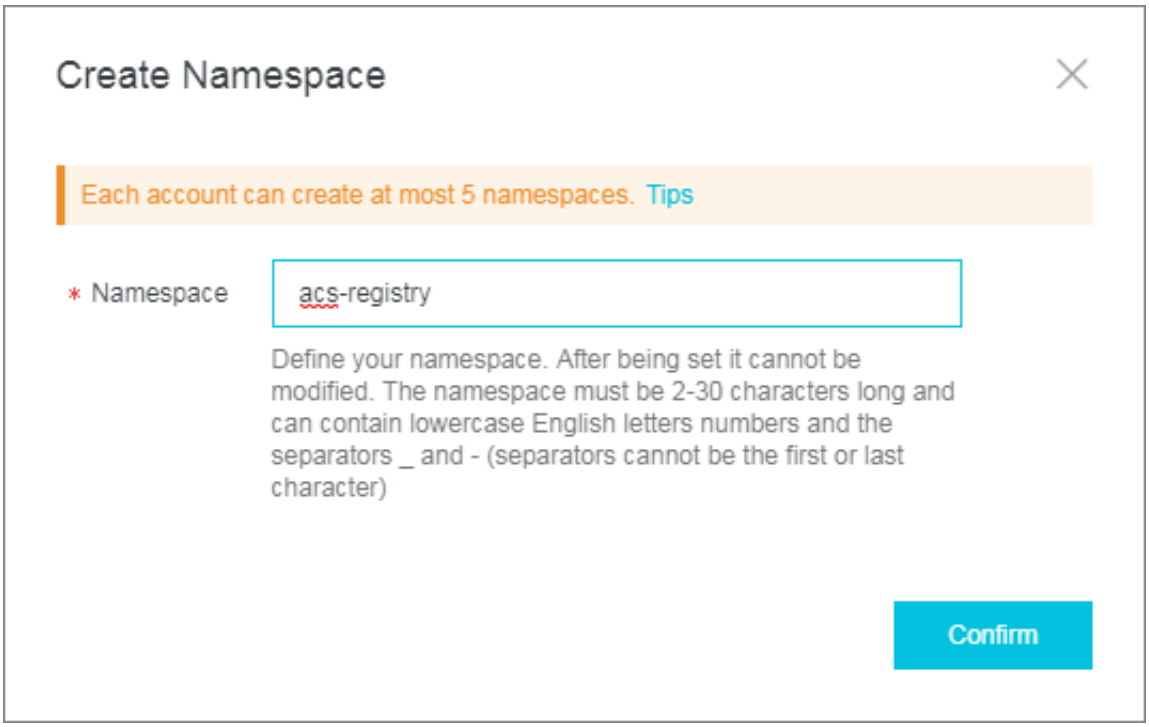
A namespace is a collection of repositories. We recommend that you group the repositories of a company or organization into a single namespace.

- **Examples of namespaces named after companies include aliyun and alibaba.**
- **Examples of namespaces named after teams or organizations include misaka-team.**

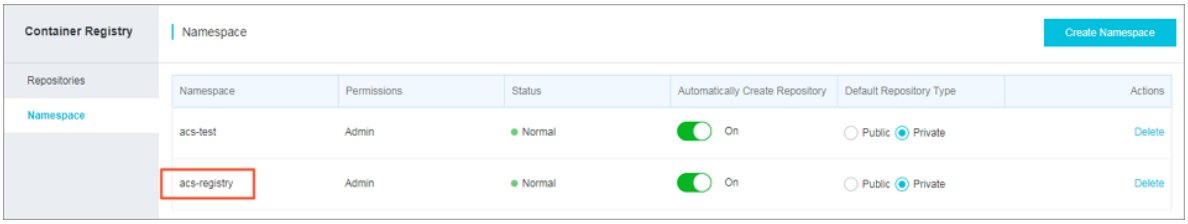
Procedure

1. *Log on to the Container Service console*.
2. **In the left-side navigation pane, choose Market > Images.**
3. **In the left-side navigation pane of the Container Registry console, click Namespaces. On the Namespaces page, click Create Namespace in the upper-**

right corner. In the Create Namespace dialog box that appears, set the parameters and click OK.



On the Namespaces page, you can view the namespace that you created and set its default repository type.



What's next

When creating a repository, you can use this namespace as a collection of repositories.

3.4.10 Use Apsara Stack Container Service for Kubernetes to release application versions in batches

Context

Note:
In the latest version of Kubernetes clusters, alicloud-application-controller is already installed by default. Only Kubernetes V1.9.3 and later are currently

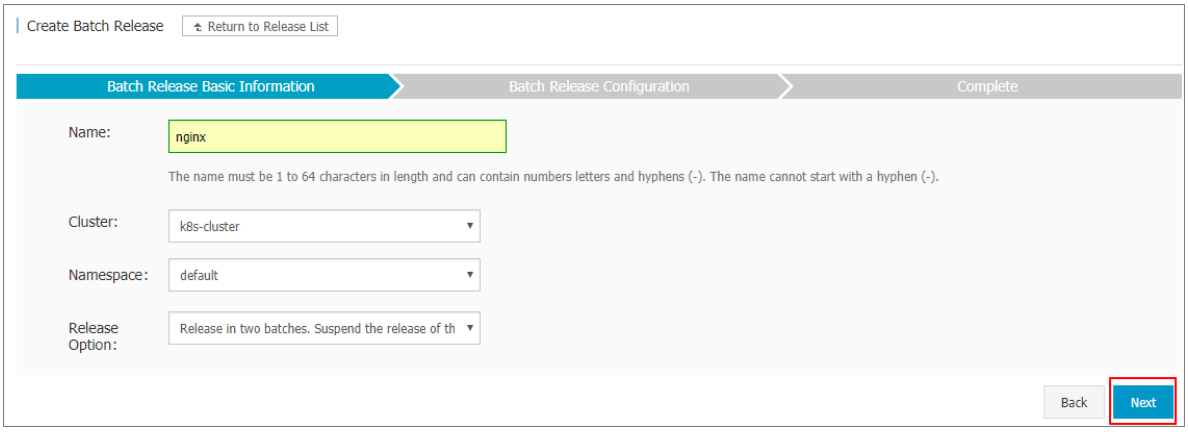
supported in Container Service. You can upgrade earlier versions of Kubernetes clusters as instructed in the Container Service console.

Procedure

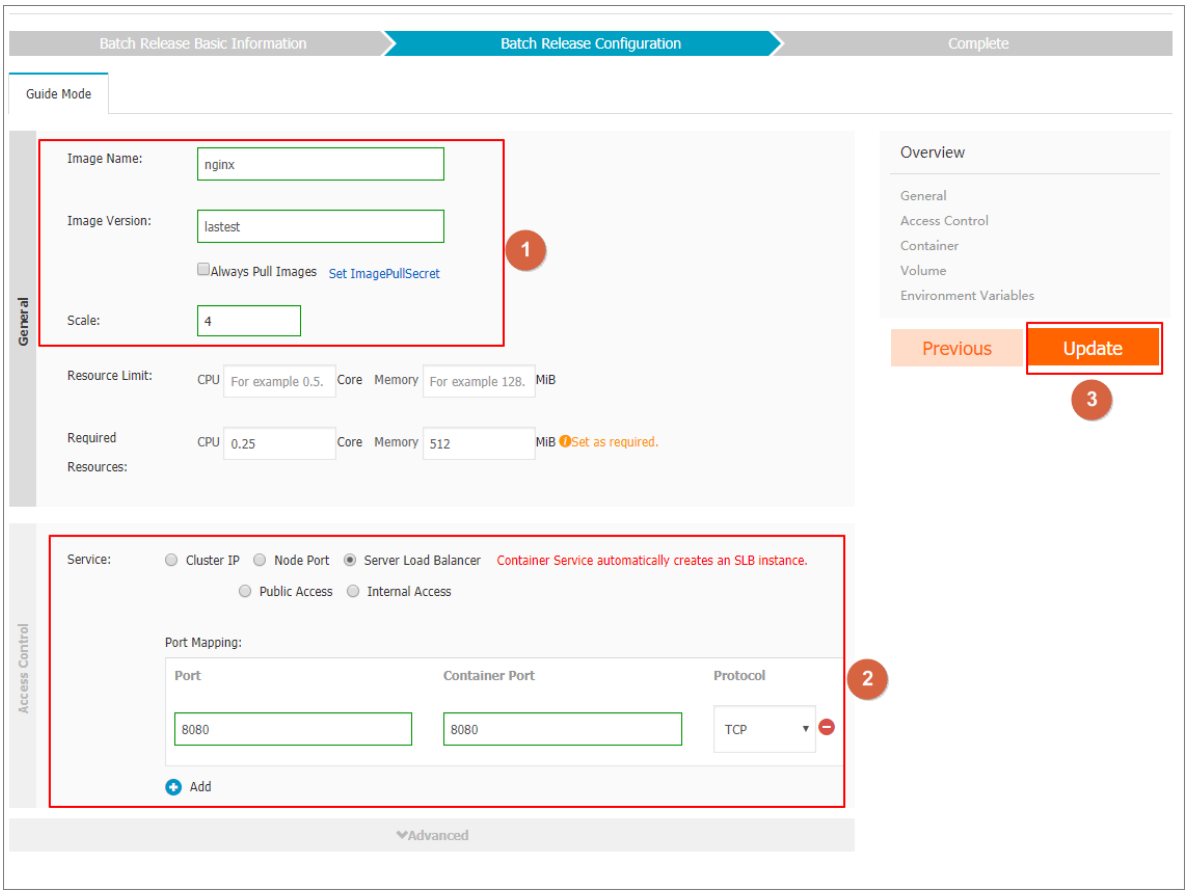
1. *Log on to the Container Service console*.
2. In the left-side navigation pane, choose Application > Releases. On the Releases page, click Create a batch release in the upper-right corner.

   Note:
   If the button is dimmed, you can refer to the upgrade link to perform the version upgrade.

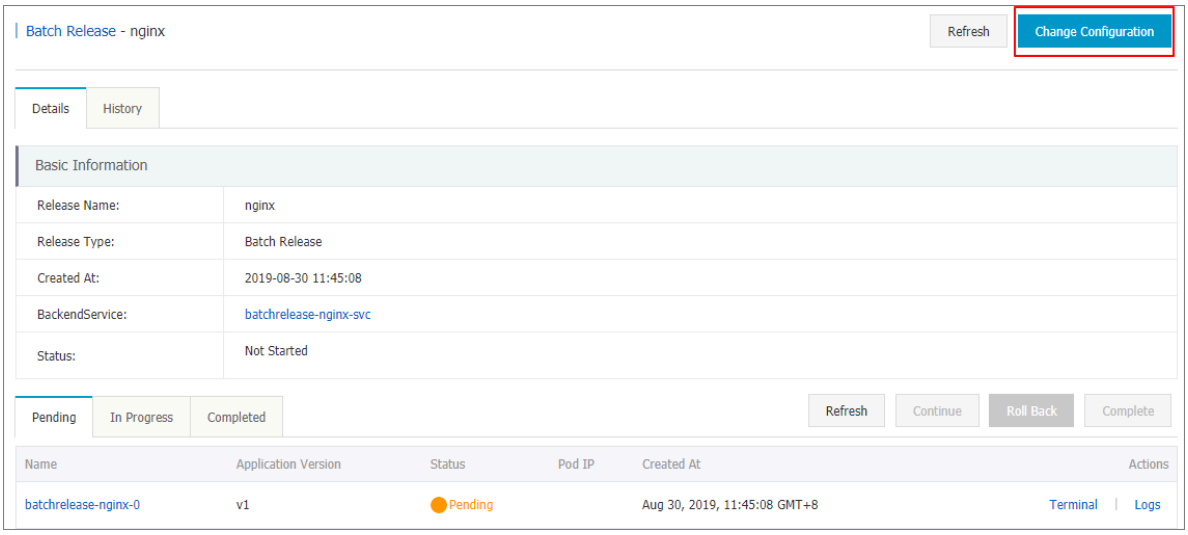
3. Set the batch release parameters, including the application name, cluster, namespace, and option. Click Next.



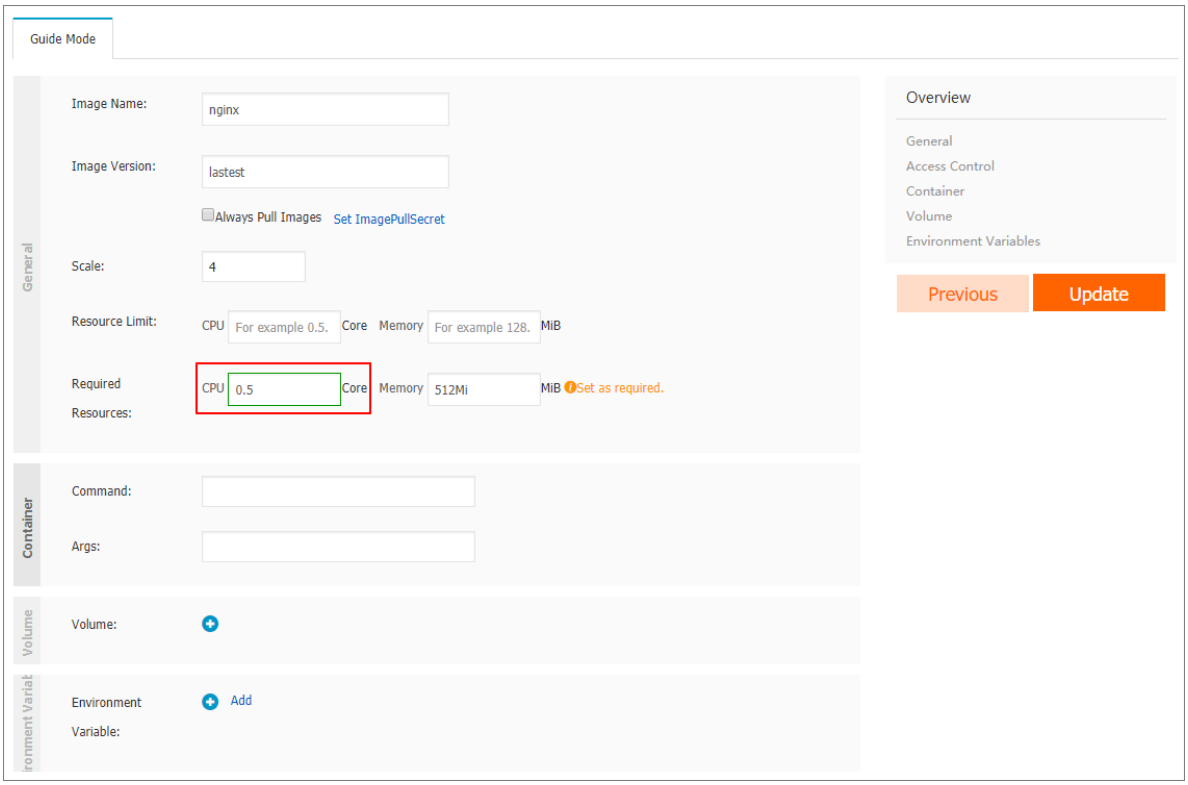
4. On the Configure Batch Release page, configure the backend pod and service, and then click Update to create an application.



5. Return to the Releases page. An application is displayed in the Not started state. Click Details in the Actions column corresponding to the application.
6. On the Details tab page, you can view more information about the application. Click Change Configuration in the upper-right corner of the page to make a batch release change.



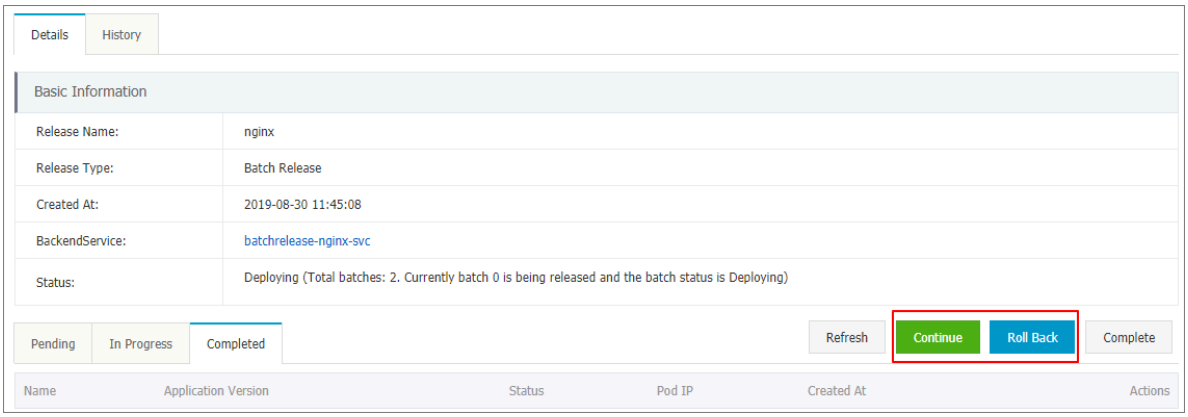
7. On the Wizard Mode page, change parameters for the new version of the application, and then click Update.



8. The Releases page appears by default, where you can view the batch release status of the application. After the first batch is deployed, click View Details.

9. The Details tab page appears. Two pods are displayed on the Not Started tab page, and another two pods are displayed on the Completed tab page. This indicates that the first batch of pods are released. Click Continue to release the

second batch of pods. Alternatively, click Roll Back to roll the application back to a previous version.



10. After the release is completed, click the History tab and roll the application back to a previous version.

What's next

You can use the batch release feature to verify your application version without traffic consumption. Batch releases consume fewer resources than blue-green releases. Batch releases can be performed on Webpages only. YAML file editors will be available later to support more complex operations.

4 Auto Scaling (ESS)

4.1 What is ESS?

Auto Scaling (ESS) is a management service that automatically adjusts the number of elastic computing resources based on your business demands and strategies.

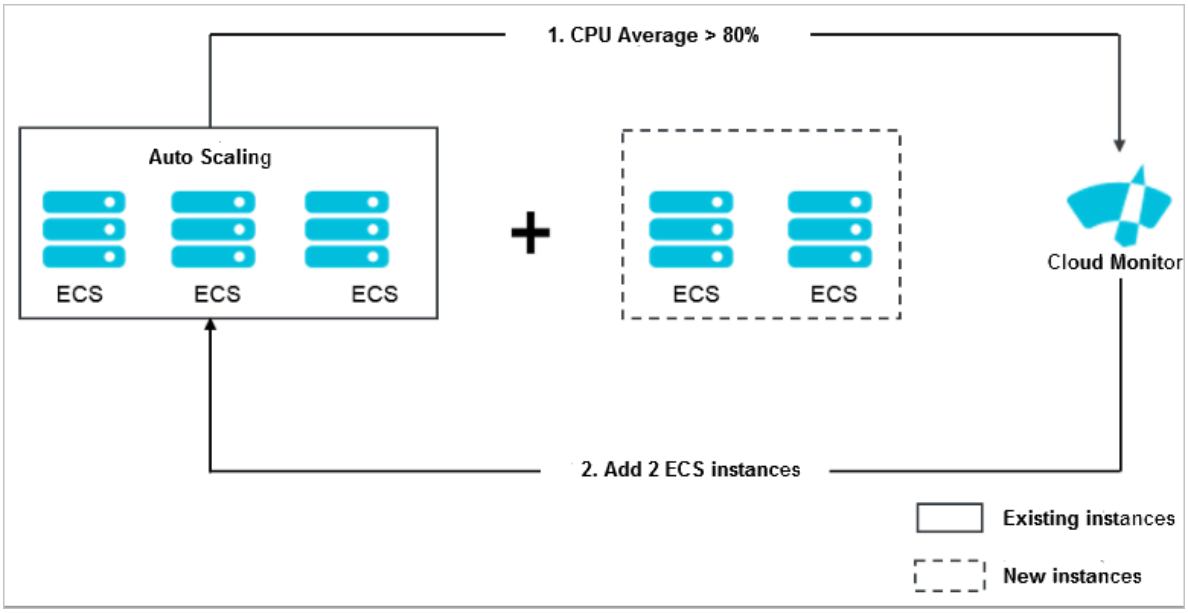
Based on user-defined scaling rules, ESS automatically adds ECS instances as business loads increase to ensure sufficient computing capabilities. When your business loads decrease, ESS automatically removes ECS instances to reduce running costs.

ESS provides the following functions:

- **Elastic scale-out**

  When business loads surge, ESS automatically increases underlying resources. This helps maintain access speed and ensure that resources are not overloaded. For example, if the CPU utilization of ECS instances exceeds 80%, ESS scales out ECS resources based on the rules you defined. During the scale-out process, ESS automatically creates and adds ECS instances to a scaling group, and adds the new instances to the SLB instance and RDS whitelist. *Figure 4-1: Elastic scale-out* **shows the process.**

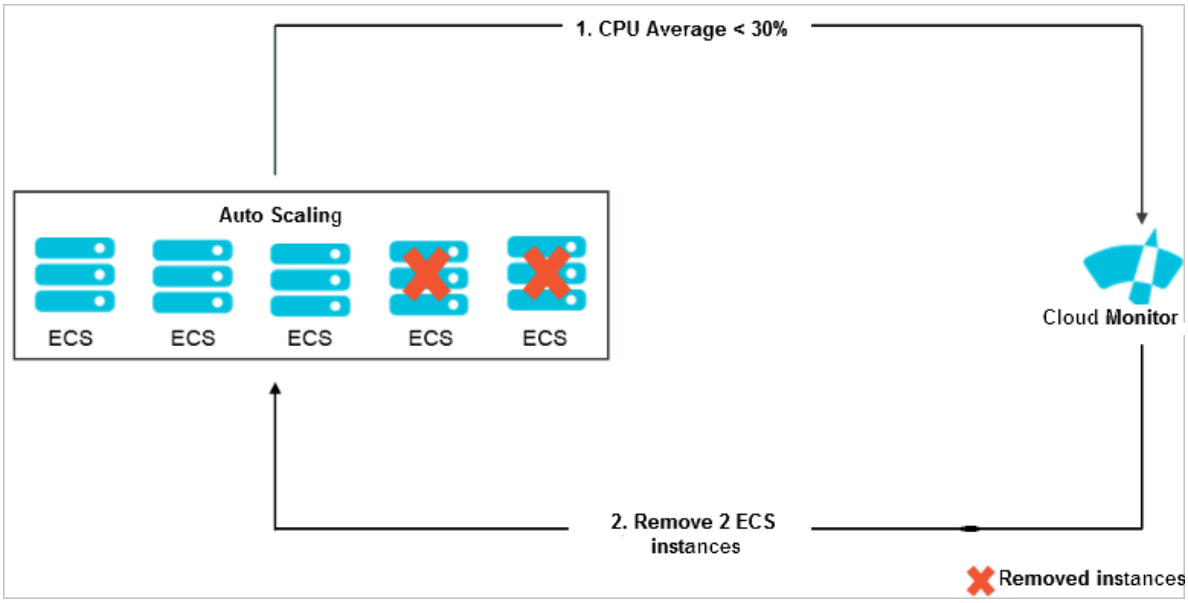
  Figure 4-1: Elastic scale-out



- **Elastic scale-in**

  When business loads decrease, ESS automatically releases underlying resources. This prevents resource wastage and helps to reduce cost. For example, if the CPU utilization of ECS instances in a scaling group falls below 30%, ESS scales in ECS resources based on the rules you defined. During the scale-in process, ESS removes the ECS instances from the scaling group, the SLB instance, and RDS whitelist. *Figure 4-2: Elastic scale-in* **shows the process.**

  Figure 4-2: Elastic scale-in



- **Elastic recovery**

  The health status of ECS instances in a scaling group is determined based on the life cycle of the instances. If an ECS instance is in an unhealthy state, ESS automatically releases the instance and creates a new one. ESS adds the new instance to the SLB instance and RDS whitelist. This process is called elastic recovery. It ensures that the number of healthy ECS instances in a scaling group will not fall below the threshold that you defined.

4.2 Usage

4.2.1 Overview

Before you use ESS, you must understand its usage limitations and take necessary precautions.

4.2.2 Precautions

This topic describes the precautions when you use ESS.

Scaling rules

During calculation and execution, a scaling rule can automatically adjust the number of ECS instances that need to be increased or decreased based on the MinSize and MaxSize values of the scaling group. For example, if the number of ECS instances to be increased that is specified by a scaling rule is 50 but MaxSize of the scaling group is 45, the scaling rule will be adjusted to increase the number of instances to a maximum of 45 instances.

Scaling activities

- **Only one scaling activity can be executed at a time in a scaling group.**
- **A scaling activity cannot be interrupted. For example, if a scaling activity to create 20 ECS instances is being executed but only five have been created, the scaling activity cannot be forcibly terminated.**
- **When a scaling activity fails to complete, the system prioritizes the integrity of the ECS instances over the scaling activity. The system will roll back the ECS instances that fail to be added or removed, but not the scaling activity. For example, if a scaling group has 20 ECS instances, out of which 19 instances are added to SLB, only the one ECS instance that failed to be added is automatically released.**

Cooldown period

- **During the cooldown period, if you manually execute a trigger task such as scaling rule or scheduled task, the task is executed immediately without being affected by the cooldown period.**
- **The cooldown period starts after the last ECS instance is added to or removed from the scaling group by a scaling activity.**

4.2.3 Manual intervention

If you manually intervene with ESS operations, ESS will process the intervention accordingly.

ESS does not prevent you from performing manual intervention, such as deleting automatically created ECS instances through the ECS console. The following table describes how ESS processes manual intervention.

Resource	Manual intervention	Processing method
ECS	A user deletes an ECS instance from a scaling group through the ECS console or open API.	ESS determines whether the ECS instance is in an unhealthy state through health check. If it is, ESS removes the instance from the scaling group. The intranet IP address of the ECS instance is not automatically deleted from the RDS access whitelist. When the number of ECS instances (Total Capacity) in the scaling group is smaller than MinSize, ESS automatically creates and adds ECS instances to the group until the number of instances is equal to MinSize.
ECS	A user revokes the ECS open API permissions granted to ESS.	ESS rejects all scaling activity requests.
SLB	A user manually removes an ECS instance from an SLB instance through the SLB console or open API.	ESS does not automatically detect this action or handle such exceptions. The ECS instance remains in the scaling group. If this instance is selected based on the removal policy during a scaling activity, it is released.
SLB	A user manually deletes an SLB instance or disables its health check function through the SLB console or open API.	ESS does not add ECS instances to scaling groups that are associated with this SLB instance. Scaling tasks can trigger scaling rules to remove ECS instances from the scaling group. ECS instances determined to be unhealthy by the health check function are also removed.

Resource	Manual intervention	Processing method
SLB	An SLB instance is unavailable because of system-related reasons.	All scaling activities fail except for instance removal tasks that are manually executed.
SLB	A user revokes the SLB open API permissions granted to ESS.	ESS rejects all scaling activity requests for scaling groups associated with SLB instances.
RDS	A user manually removes the IP address of an ECS instance from an RDS whitelist through the RDS console or open API.	ESS does not automatically detect this action or handle such exceptions. The ECS instance remains in the scaling group. If this instance is selected based on the removal policy during a scaling activity, it is released.
RDS	A user manually deletes an RDS instance through the RDS console or open API.	ESS does not add ECS instances that are associated with this RDS instance to scaling groups. Scaling tasks can trigger scaling rules to remove ECS instances from the scaling group. ECS instances determined to be unhealthy by the health check function are also removed.
RDS	An RDS instance is unavailable because of system-related reasons.	All scaling activities fail except for instance removal tasks that are manually executed.
RDS	A user revokes the RDS open API permissions granted to ESS.	ESS rejects all scaling activity requests for the scaling groups associated with RDS instances.

4.2.4 Quantity limits

Before you use ESS, you need to understand the quantity limits of ESS.

You can only create a limited number of scaling groups, scaling configurations, scaling rules, scaling ECS instances, and scheduled tasks.

Table 4-1: Quantity limits

Item	Description
Scaling group	You can create a maximum of 20 scaling groups.

Item	Description
Scaling configuration	You can create a maximum of 10 scaling configurations in a scaling group.
Scaling rule	You can create a maximum of 10 scaling rules in a scaling group.
ECS instances for scaling	You can configure a maximum of 100 ECS instances for automatic scaling. **Note:** This limit applies to the ECS instances that are automatically created, but does not apply to manually added ones.
Scheduled task	You can create a maximum of 20 scheduled tasks.

4.2.5 Scaling group statuses

Before you manage a scaling group, you need to understand the scaling group statuses.

A scaling group can be in Active, Inactive, or Deleting state. *Table 4-2: Scaling group statuses* **describes the details.**

Table 4-2: Scaling group statuses

Status	Status in open API
Creating	Inactive
Created	Inactive
Enabling	Inactive
Running	Active
Disabling	Inactive
Stopped	Inactive
Deleting	Deleting

4.2.6 Scaling activity process

Before you use ESS, you need to understand the processes related to the scaling activity.

Automatic scaling of a scaling group

Automatic scale-out

1. **Check the health status and other prerequisites for scaling.**
2. **Allocate the activity ID and execute the scaling activity.**
3. **Create ECS instances.**
4. **Modify Total Capacity.**
5. **Allocate IP addresses to the created ECS instances.**
6. **Add ECS instances to the RDS whitelist.**
7. **Start ECS instances.**
8. **Associate the ECS instances to an SLB instance and set the weight to the `SLB weight` value when the scaling configuration is created.**
9. **The scaling activity completes, and the cooldown period starts.**

Automatic scale-in

1. **Check the health status and other prerequisites for scaling.**
2. **Allocate the activity ID and execute the scaling activity.**
3. **Remove ECS instances from the SLB instance.**
4. **Stop ECS instances.**
5. **Remove ECS instances from the RDS whitelist.**
6. **Release ECS instances.**
7. **Modify Total Capacity.**
8. **The scaling activity completes, and the cooldown period starts.**

Manually adding or removing existing ECS instances

Manually adding

1. **Check the health status and other prerequisites for scaling, and check the status and type of ECS instances.**
2. **Allocate the activity ID and execute the scaling activity.**
3. **Add ECS instances.**
4. **Modify Total Capacity.**

5. Add ECS instances to the RDS whitelist.
6. Associate ECS instances to an SLB instance and set the weights to the `SLB weight` value of the active scaling configuration.

> **Note:**
> When you need to manually add an instance, the instance type must be the same as that specified in the active scaling configuration of the scaling group. Therefore, you must set the weight to the `SLB weight` value specified in the scaling configuration.

7. The scaling activity completes, and the cooldown period starts.

Manual removal

1. Check the health status and boundary conditions of a scaling group.
2. Allocate the activity ID and execute the scaling activity.
3. SLB stops forwarding traffic to ECS instances.
4. Remove ECS instance from SLB after 60 seconds.
5. Remove ECS instances from the RDS whitelist.
6. Modify Total Capacity.
7. Remove ECS instances from the scaling group.
8. The scaling activity completes, and the cooldown period starts.

> **Note:**
> The life cycle of a scaling activity starts at checking the health status and other prerequisites for scaling. and ends at starting the cooldown time.

4.2.7 Removal of unhealthy ECS instances

Before you use ESS, you need to read information about the removal of unhealthy ECS instances.

After an ECS instance has been successfully added to a scaling group, ESS periodically scans its status. If the ECS instance is not in Running state, ESS removes the ECS instance from the scaling group.

- If an ECS instance is created automatically, ESS immediately removes and releases it.
- If the ECS instance is added manually by a user, ESS immediately removes it, but does not stop or release it.

The MinSize attribute of a scaling group does not limit the removal of unhealthy ECS instances. That is, the total number of ECS instances can fall below MinSize after the removal. ESS automatically creates ECS instances based on the difference between the actual instance number and MinSize to ensure the total number is equal to MinSize.

4.2.8 Instance rollback after a scaling activity failure

Before you use ESS, you need to understand the mechanism of instance rollback after a failed scaling activity.

When a scaling activity fails to complete, the system prioritizes the integrity of the ECS instances over the scaling activity. The system will roll back the ECS instances that fail to be added or removed, but not the scaling activity. That is, the system rolls back ECS instances, not the scaling activity.

Example

If a scaling group has created 20 ECS instances, out of which 19 instances are added to SLB, only the one ECS instance that failed to be added is automatically released.

4.2.9 Instance life cycle management

Before you use ESS, you need to understand the concepts related to instance life cycles.

ECS instances in a scaling group can be created automatically or added manually.

Automatically created ECS instances

ECS instances are automatically created by ESS based on user-defined scaling configurations and rules.

ESS manages the entire life cycle of this type of ECS instances. ESS creates this type of ECS instances during scale-out, and stops and release them during scale-in.

Manually added ECS instances

ECS instances are manually added to a scaling group.

ESS does not manage the entire life cycle of this type of ECS instances. Such instances are not created by ESS, but are manually added by a user to a scaling group. When the ECS instances are removed from a scaling group manually or as the result of an automatic scale-in, ESS removes the instances but does not stop or release them.

Instance status

An ECS instance in a scaling group undergoes the following status during its life cycle:

- **Pending: The ECS instance is being added to a scaling group. For example, ESS is creating the instance or adding it to an SLB instance or RDS whitelist.**
- **In Service: The ECS instance has been successfully added to a scaling group and is providing services normally.**
- **Removing: The ECS instance is being removed from a scaling group.**

Instance health statuses

An ECS instance in a scaling group has the following health statuses:

- **Healthy**
- **Unhealthy**

If an ECS instance is not in Running state, it is considered as an unhealthy instance. ESS automatically removes unhealthy ECS instances from a scaling group.

- **ECS instances that are automatically created are stopped and released by ESS.**
- **ECS instances that are manually added are not stopped and released by ESS.**

4.3 Quick start

4.3.1 Overview

This topic describes how to quickly create scaling groups, configurations, and rules. It is designed to guide you through the process of automatic scaling creation.

Follow these steps:

1. *Create a scaling group*

   Configure information such as MinSize and MaxSize attributes of scaling resources, as well as SLB and RDS instances to be associated with a scaling group .

2. *Create a scaling configuration*

   Configure ECS instance configurations for automatic scaling, such as Image ID and Instance Type.

3. *Enable a scaling group*

   Enable the scaling group created in step 2.

4. *Create a scaling rule*

   Create a scaling rule based on actual conditions. ESS executes scaling based on the specified rule, such as adding N ECS instances.

5. *Create a scheduled task*

   Create a scheduled task based on actual conditions. ESS executes scaling rules at a specified point in time. For example, you can create a scheduled task to execute the scaling rule created in step 4 at 12:00.

4.3.2 Log on to the Auto Scaling console

This topic describes how to log on to the Auto Scaling console.

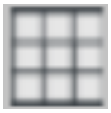
Prerequisites

- **Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://`IP address or domain name of the Apsara Stack console`/manage.**
- **We recommend that you use the Chrome browser.**

Procedure

1. **Open your browser.**
2. **In the address bar, enter the access address of the Apsara Stack console in the format of http://`IP address or domain name of the Apsara Stack console`/manage, and then press Enter.**
3. **Enter the correct username and password.**
   - **The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.**
   - **You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers**

(0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).

4. Click LOGIN to go to the Dashboard page.
5. In the top navigation bar, choose  > Compute, Storage & Networking > Auto Scaling.

4.3.3 Create a scaling group

You must create scaling groups before you can use the services provided by ESS.

Prerequisites

- The scaling groups must be in the same region as the SLB and RDS instances that they will be associated with.
- A maximum of 20 scaling groups can be created by a user.

Procedure

1. *Log on to the Auto Scaling console*.
2. Navigate to the Scaling Groups page, and click Create Scaling Group. On the page that appears, configure relevant parameters.

   Table 4-3: Parameters for creating a scaling group describes the configurations of each parameter.

   Table 4-3: Parameters for creating a scaling group

Type	Parameter	Description
Region	Region	Required. The region to which the scaling group belongs.
Basic Settings	Department	Required. The department to which an instance belongs.
	Project	Required. The project to which an instance belongs.
	Scaling Group Name	Required. A scaling group name must be 2 to 40 characters in length. It can contain numbers, uppercase letters , lowercase letters, underscores (_), hyphens (-), and periods (.). It must start with a number or letter.

Type	Parameter	Description
	Minimum Instances	Required. The minimum number of instances that a scaling group must contain to ensure availability. After you have completed this scaling group configuration, the system creates a group containing the number of instances as specified here. Value range: 0–100.
	Maximum Instances	Required. The maximum number of instances that a scaling group can contain, to control costs. Value range: 0–100.
	Default Cooldown Time	Required. The cooldown period for a scaling group. After a scaling activity has been successfully executed and the last ECS instance is added to or removed from the group, the cooldown period starts immediately. During the cooldown period, this scaling group cannot execute any new scaling activities. The value must be an integer with a minimum value of 0. Value range: 0–86400.
	Removal Policy	Optional. This policy is used to filter and remove ECS instances from a scaling group using multiple filtering conditions.
Network type	VPC	Virtual Private Cloud (VPC): VPC helps you build an isolated network environment in Apsara Stack. You can customize routing tables, IP address segments, and gateways in a VPC. We recommend that you set Network Type to VPC to improve security. Before you set Network Type to VPC, ensure that you have created a VPC and VSwitch. For more information, see Create a VPC and Create a VSwitch in *VPC User Guide*.

Type	Parameter	Description
	Classic Network	Cloud services in a classic network are not isolated. Unauthorized access to cloud services is blocked only by the security group or whitelist policy.
Whitelist Configuration	SLB Instances	Optional. The SLB instance to be associated with the scaling group. If an SLB instance is specified for a scaling group, the scaling group automatically adds its ECS instances to the specified SLB instance. Note: · The specified SLB instance must be active. · An ECS instance that is added for load balancing has a default weight of 50.
	Databases	· Optional. If an RDS instance is specified for a scaling group, the scaling group automatically adds the intranet IP addresses of its ECS instances to the whitelist of the specified RDS instance. Note: - The specified RDS instance must be in the `Running` state. - The number of IP addresses in the whitelist of the specified RDS instance cannot exceed the whitelist upper limit. - A scaling group does not take effect immediately after creation. It must be enabled before scaling rules can be triggered and scaling activities can be executed.

3. After you complete the parameter configurations, click Create.

4.3.4 Create a scaling configuration

You can create a scaling configuration to customize the specifications of the ECS instances that are to be automatically added to a scaling group.

Prerequisites

Ensure that at least one security group is available. If you do not have any security groups, you need to create a security group. For more information, see *the Create security groups section of* **ECS User Guide.**

Procedure

1. *Log on to the Auto Scaling console*.
2. **On the Scaling Groups tab, click the ID of the scaling group for which you want to create a scaling configuration.**
3. **On the displayed page, click the Configure Scaling tab and then click Create Scaling Configuration. Configure parameters on the displayed page.**

   For more information about the parameters and descriptions, see *Table 4-4: Parameters for creating a scaling configuration*.

   Table 4-4: Parameters for creating a scaling configuration

Category	Parameter	Description
Basic Settings	Configuration Name	The name of the scaling configuration. The name must be 2 to 40 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). It must start with a digit or letter.
	Security Groups	The security group to which an instance belongs.
Instance	Instance Series	The default value is Series 2.
	I/O Optimized	The default value is I/O-Optimized Instances.
	Instance Specification	The instance type that you need.
Network	Public Bandwidth	The method used for calculate billing for public bandwidth. The values include Pay By Traffic and By Fixed Bandwidth.

Category	Parameter	Description
	Bandwidth	The bandwidth that you need. You can adjust the slider to set the bandwidth.
Images	Image Type	· Click Public Image to select an operating system and version as required. · If you need to enable features such as automatic startup of Web servers and automatic downloads of code and scripts, select Custom Image.
Storage	System Disk	The type and size of the cloud disk that is used to install the system image.
	Data Disk	The data disk to be added. You can select the type, size, and mount point of the data disk. In the current Auto Scaling version, you can add up to four data disks to each ECS instance.

4. After you complete the parameter configurations, click Create.

4.3.5 Enable a scaling group

Before you use a scaling group, you must manually enable the group.

Prerequisites

- The scaling group is in the `Inactive` state.
- The scaling group has an active scaling configuration.
- A single scaling group can only have one `Active` scaling configuration at a time.

Procedure

1. *Log on to the Auto Scaling console*.
2. On the Scaling Groups tab page, locate the scaling group that you want to enable, click the icon in the `Actions` column, and choose Enable from the shortcut menu.
3. In the message that appears, click `OK`.

Result

The scaling group status will change from `Inactive` to `Active`.

4.3.6 Create a scaling rule

This topic describes how to create a scaling rule after you create a scaling group.

Context

- **You can create up to 10 scaling rules for each scaling group.**
- **After a scaling rule is executed, the number of ECS instances in the scaling group may not meet the Minimum Instances or Maximum Instances requirement. In this case, Auto Scaling will automatically adjust the number of the ECS instances to meet the maximum or minimum instances requirement.**
- **After a scaling rule is created, a unique identifier is generated. The unique identifier can be used in the following API operations:**
  - **Scaling rule execution operation: You can specify the identifier in the ScalingRuleAri parameter of the operation to manually execute the scaling rule.**
  - **Scheduled task creation operation: You can specify the identifier in the ScheduledAction parameter to execute the scaling rule at the scheduled time.**

Procedure

1. *Log on to the Auto Scaling console*.
2. **Click the Scaling Groups tab, find the target scaling group, and click the instance ID. You are redirected to the Basic Information page.**
3. **Click the Scaling Rules tab, and then click Create Scaling Rule in the upper-right corner.**

4. **In the Create Scaling Rule dialog box that appears, set the parameters.**

   For more information about the parameters and descriptions, see *Table 4-5: Parameters and descriptions*.

   Table 4-5: Parameters and descriptions

Parameter	Description
Rule Name	**Specify a name for the scaling rule. The name must be from 2 to 40 characters in length and can contain numbers, underscores (_), hyphens (-), or periods (.). It must start with a number, letter (case -insensitive), or Chinese character.**
Rule Action	**Select** `Change To`**,** `Add`**, or** `Remove` **from the drop-down list, and enter a number in the text box to specify the number of ECS instances.**
Cooldown Time	**Specify the cooldown period.** **Note:** **If this parameter is left empty, the cooldown period of the scaling group is selected by default.**

5. **Click OK.**

4.3.7 Create a scheduled task

This topic describes how to create a scheduled task in the Auto Scaling console.

Prerequisites

- **To create a scheduled task, set the parameters as required. You can create up to 20 scheduled tasks.**
- **If multiple tasks are scheduled to run at the same time point, the most recently created scheduled task is executed.**

Procedure

1. *Log on to the Auto Scaling console*.

2. **Click the Scheduled Tasks tab, click Create Scheduled Task in the upper-right corner, and then set the parameters.**

   For more information about the parameters and descriptions, see *Table 4-6: Parameters for scheduled task configuration*.

   Table 4-6: Parameters for scheduled task configuration

Category	Parameter	Description
Region	Region	Required. The region is automatically selected by the system. You do not need to specify a region.
Basic Settings	Department	Required. Specify the department to which the scheduled task belongs.
	Project	Required. Select the project to which the scheduled task belongs.
	Task Name	Required. The name must be 2 to 40 characters in length and can contain digits, letters, underscores (_), hyphens (-), and periods (.). It must start with a digit or letter.
	Description	Specify description for the task. The description must be at least two characters in length.
	Execution Time	Required. Specify the time to execute the task.
	Scaling Rules	Required. Select a scaling group and scaling rule for the task.
	Retry Expiry Time (Seconds)	Optional. Specify the time period after which all retry attempts will stop.
Recurrence	Set	Specify this parameter to enable or disable periodical task execution. `Not Set` is selected by default. If you select `Set`, you must set the `Recurrence` and `End Time` parameters.

3. **Click Create.**

4.4 Scaling group

4.4.1 Overview

A scaling group is a group of ECS instances that is dynamically scaled based on the configured scenario. You can specify the maximum and minimum number of ECS instances in a scaling group, as well as SLB and RDS instances associated with the group.

4.4.2 Query a scaling group

You can query created scaling groups and their related information in the ESS console.

Procedure

1. *Log on to the Auto Scaling console*.
2. **On the Scaling Groups tab page, set filtering conditions such as** `Department`, `Region`, **or** `Instance Name`, **and click Search.**

   Note:

   - **Click Instance Name, you can select other filtering conditions from the drop-down list, including** `Instance ID`, `Status`, **and** `Project Name`.
   - **On the Scaling Groups tab page, you can click an instance ID to go to the Basic Information page.**
3. **Click the icon in the** `Actions` **column corresponding to an instance, and choose View Details from the shortcut menu. The Basic Information page appears.**

4.4.3 Edit a scaling group

This topic describes how to change the information of a scaling group in the Auto Scaling console.

Context

If the number of ECS instances in a scaling group drops below the MinSize value or exceeds the MaxSize value, ESS automatically adds instances to or removes instances from the group. This ensures that the number of instances always meet the maximum or minimum instances requirement.

Procedure

1. *Log on to the Auto Scaling console*.
2. Click the Scaling Groups tab and find the target scaling group. In the `Actions` column, click the icon and select Change.
3. In the Change Scaling Group Information dialog box that appears, you can change the parameters as needed.
4. Click Change.

4.4.4 Disable a scaling group

You can disable scaling groups in the ESS console.

Prerequisites

- You can disable a scaling group only when it is not executing any scaling activity.
- You can successfully disable a scaling group only when it is in the `Active` state.

Procedure

1. *Log on to the Auto Scaling console*.
2. On the Scaling Groups tab page, locate the scaling group to be disabled, click the icon in the `Actions` column, and choose Disable from the shortcut menu.
3. In the message that appears, click `OK`.

Result

The scaling group status changes from `Active` to `Inactive`.

4.4.5 Delete a scaling group

You can delete scaling groups that you no longer use in the ESS console.

Context

- If you delete a scaling group through the ESS console, the group is forcibly deleted.
- Deleting a scaling group also deletes all scaling configurations, rules, activities, and requests related to the group.
- Deleting a scaling group does not delete scheduled tasks, SLB instances, or RDS instances related to the group.

Procedure

1. *Log on to the Auto Scaling console*.
2. **On the Scaling Groups tab page, locate the scaling group to be deleted, click the icon in the `Actions` column, and choose Delete from the shortcut menu.**
3. **In the message that appears, click `OK`.**

4.4.6 Query ECS instances

You can query ECS instances in the ESS console.

Procedure

1. *Log on to the Auto Scaling console*.
2. **On the Scaling Groups tab page, locate the scaling group whose ECS instance you want to query, click the icon in the `Actions` column, and choose View Details from the shortcut menu.**

3. **Click the ECS Instances tab to view instance details.**

   - **Query ECS instances in a scaling group**

     There are two types of ECS instances in a scaling group:

     - **Automatically created: ECS instances that are created automatically based on scaling configurations and rules.**
     - **Manually added: ECS instances that are added manually to a scaling group by a user.**

   - **Life cycle of ECS instances in a scaling group:**

     - `Adding`**: An ECS instance is being added to a scaling group. For example, the instance is being created or added to an SLB or RDS whitelist.**
     - `In Service`**: An ECS instance has been successfully added to a scaling group and is providing services normally.**
     - `Removing`**: An ECS instance is being removed from a scaling group.**

   - **ECS health status**

     Health status of ECS instances:

     - **Healthy**
     - **Unhealthy**

     ESS automatically removes unhealthy ECS instances from a scaling group. ECS instances that are automatically created are stopped and released by ESS. ECS instances that are manually added are not stopped or released by ESS.

4.5 Scaling configuration

4.5.1 Overview

Scaling configurations specify the specifications of ECS instances used for automatic scaling. When automatically adding ECS instances to a scaling group, ESS will create ECS instances based on the scaling configurations.

4.5.2 Query a scaling configuration

You can query created scaling configurations and their related information in the ESS console.

Procedure

1. *Log on to the Auto Scaling console*.
2. On the Scaling Groups tab page, locate the scaling group of which you want to query the scaling configuration, and click a configuration in the `Scaling Configuration` column corresponding to the group. The Scaling Configurations tab page appears.
3. On the Scaling Configurations tab page, click the icon in the `Actions` column and choose View Details from the shortcut menu.

> **Note:**
> You can also click a scaling configuration name on the Scaling Configurations tab page to view configuration details.

4.6 Scaling rule

4.6.1 Overview

Scaling rules define specific scaling actions executed by ESS, such as scaling in and out ECS instances.

4.6.2 Query a scaling rule

You can query created scaling rules and their related information in the ESS console.

Procedure

1. *Log on to the Auto Scaling console*.
2. On the Scaling Groups tab page, locate the scaling group for which you want to query a scaling rule, and click the group ID. The Basic Information page appears.
3. Click the Scaling Rules tab, locate the scaling rule you are searching for, and click the icon in the `Actions` column.
4. Choose View Details from the shortcut menu. The scaling rule details are displayed.

4.6.3 Edit a scaling rule

This topic describes how to edit a scaling rule in the Auto Scaling console.

Procedure

1. *Log on to the Auto Scaling console*.
2. Click the Scaling Groups tab, find the target scaling group, and click the instance ID. You are redirected to the Basic Information page.
3. Click the Scaling Rules tab and find the target scaling rule. In the `Actions` column, click the icon and select Change.
4.
5. You can change the `Rule Name`, `Rule Action`, and `Cooldown Time` parameters.

   Note:

   - `Rule Name`: The name must be from 2 to 40 characters in length and can contain numbers, underscores (_), hyphens (-), and periods (.). It must start with a number, letter (case-insensitive), or Chinese character.
   - `Cooldown Time`: If this parameter is left empty, the cooldown period of the scaling group will be selected by default.
6. Click OK.

4.6.4 Delete a scaling rule

You can delete scaling rules that you no longer use in the ESS console.

Procedure

1. *Log on to the Auto Scaling console*.
2. On the Scaling Groups tab page, locate the scaling group that contains the scaling rule to be deleted, and click the group ID. The Basic Information page appears.
3. Click the Scaling Rules tab, locate the scaling rule to be deleted, click the icon in the `Actions` column, and choose Delete from the shortcut menu.
4. In the Delete Scaling Rule message that appears, click OK.

4.7 Trigger tasks

4.7.1 Overview

In the ESS console, you can perform automatic scaling by manually executing scaling rules or adding ECS instances.

4.7.2 Manually execute a scaling rule

This topic describes how to manually execute a scaling rule.

Prerequisites

If you need to execute a scaling rule, note the following limits:

- **The status of the scaling group including the scaling rule must be** `Active`**.**
- **The scaling group including the scaling rule is not executing any scaling activity.**
- **An Apsara Stack tenant account can automatically scale up to a maximum of 1, 000 ECS instances across all scaling groups in all regions. This limit applies to the ECS instances that are automatically created, but does not apply to manually added ECS instances.**
- **ESS automatically scales ECS instances to ensure that the actual number of instance does not exceed the limits.**

Procedure

1. *Log on to the Auto Scaling console*.
2. **On the Scaling Groups tab page, locate the scaling group for which you want to execute a scaling rule, and click the group ID. The Basic Information page appears.**
3. **Click the Scaling Rules tab, locate the scaling rule to be executed, click the**  **icon in the** `Actions` **column, and choose Execute from the shortcut menu.**

4.7.3 Add an ECS instance

You can add ECS instances to a specific scaling group in the Auto Scaling console.

Prerequisites

Only ECS instances that meet the following conditions can be added to the scaling group:

- **It must be in the same region as the scaling group.**

- The instance type must be the same as that specified in the active scaling configuration.
- It must be in the `Running` state.
- It cannot belong to any other scaling group at the same time.
- If the network type of the scaling group is VPC, only instances belonging to the same VPC as the scaling group can be added.

Before adding an ECS instance, you must ensure that:

- The scaling group must be in the `Active` state.
- No scaling activities are being executed in the scaling group.

> **Note:**
>
> - If no scaling activities are being executed in the scaling group, you can immediately remove ECS instances and do not need to wait after the cooldown period.
> - If the number of instances will exceed MaxSize after manual addition, the add operation fails.
> - Manually added ECS instances are not associated with the active scaling configuration in the scaling group.

Procedure

1. *Log on to the Auto Scaling console*.
2. On the Scaling Groups tab, find the scaling group to which you want to add ECS instances and click the Group ID. The Basic Information page appears.
3. Click the ECS Instances tab and click Add Existing Instances.
4. In the dialog box that appears, select the target instances in the right-side `Available Instances` list, click the  icon to add instances to the left-side `Selected Instances` list, and click OK.

> **Note:**
>
> - You can click All to select all instances in a list.
> - Click the  icon to remove selected instances.

4.7.4 Remove an ECS instance

You can remove ECS instances from a scaling group in the ESS console.

Prerequisites

- **When an automatically created ECS instance is removed from a scaling group, the instance is stopped and released.**
- **When a manually added ECS instance is removed from a scaling group, the instance is not stopped or released.**

Prerequisites to remove an ECS instance:

- **The scaling group must be** `Active`**.**
- **No scaling activities are being executed in the scaling group.**

> **Note:**
>
> - **If no scaling activities are being executed in the scaling group, you can immediately remove ECS instances and do not need to wait after the cooldown period.**
> - **If a manual remove operation would cause the number of instances be less than MinSize, the remove operation fails.**

Procedure

1. *Log on to the Auto Scaling console*.
2. **On the Scaling Groups tab page, locate the scaling group from which you want to remove ECS instances, and click the group ID. The Basic Information page appears.**
3. **Click the ECS Instances tab. Click Auto Created or Manually Added on the tab page.**
4. **Locate the instance to be removed, click the icon in the** `Actions` **column, and choose Remove and Release from the shortcut menu.**
5. **In the message that appears, click OK.**

4.8 Scheduled tasks

4.8.1 Overview

If a scaling group is disabled or executing a scaling activity, a scheduled task fails to execute a scaling rule. The scheduled task is automatically retried within LaunchExpirationTime. After LaunchExpirationTime expires, the task is abandoned. If multiple tasks in the same group are scheduled at similar points in time, the earliest task executes its scaling activity first. A scaling group can execute only one scaling activity at a time. Other tasks attempt to execute the rule within LaunchExpirationTime. If a scaling activity is completed within LaunchExpirationTime, the completed activity will trigger the next scheduled scaling rule and execute the scaling activity.

4.8.2 Query a scheduled task

In the ESS console, you can query created scaling rules and related information.

Procedure

1. *Log on to the Auto Scaling console*.
2. **Click the Scheduled Task tab, locate a task to be queried, and click the icon in the `Actions` column.**
3. **Choose View Details from the shortcut menu. The task details are displayed.**

4.8.3 Edit a scheduled task

This topic describes how to change the information about a scheduled task in the Auto Scaling console.

Procedure

1. *Log on to the Auto Scaling console*.
2. **Click the Scheduled Tasks tab and find the target scheduled task. In the Actions column, click the icon and select Change.**

3. **In the Change Scheduled Task dialog box that appears, modify the information about the task.**

   Table 4-7: Parameters and descriptions

Parameter	Description
Task Name	You can rename the scheduled task. The name must be from 2 to 40 characters in length and can contain numbers, underscores (_), hyphens (-), or periods (.). It must start with a number, letter (case-insensitive), or Chinese character.
Description	You can add a description to the task. The description must be at least two characters in length.
Execution Time	Specify the time to execute the task.
Scaling Rules	Select scaling rules for the task.
Retry Timeout	Specify the time period after which all retry attempts will stop.
Schedule Settings	Set is selected by default.
Recurrence	Specify an interval to periodically run the scheduled task.
End Time	Specify the end time of the periodical execution process.

4. **Click OK.**

4.8.4 Stop a scheduled task

This topic describes how to stop a scheduled task in the Auto Scaling console.

Prerequisites

The scheduled task is in the Running state.

Procedure

1. *Log on to the Auto Scaling console*.
2. **Click the Scheduled Tasks tab and find the target scheduled task. In the Actions column, click the icon and select Stop.**
3. **Click OK.**

4.8.5 Start a scheduled task

This topic describes how to start a scheduled task in the Auto Scaling console.

Prerequisites

The scheduled task is in the Stopped state.

Procedure

1. *Log on to the Auto Scaling console*.
2. **Click the Scheduled Tasks tab and find the target scheduled task. In the Actions column, click the icon and select Start.**
3. **Click OK.**

4.8.6 Delete a scheduled task

You can delete scheduled tasks that you no longer use in the ESS console.

Procedure

1. *Log on to the Auto Scaling console*.
2. **Click the Scheduled Task tab, locate a scheduled task to be deleted, click the icon in the `Actions` column, and choose Delete from the shortcut menu.**
3. **In the message that appears, click OK.**

4.9 Monitoring tasks

4.9.1 Overview

Auto Scaling monitoring tasks integrate the features of Auto Scaling and CloudMonitor. You can create monitoring tasks to help you dynamically manage your scaling groups. After you associate monitoring tasks with monitoring metrics of CloudMonitor, these monitoring tasks can perform specified scaling activities when certain scaling rules are triggered. This helps you adjust computing resources based on your actual business needs.

4.9.2 Create a monitoring task

This topic describes how to create a monitoring task in the Auto Scaling console.

Procedure

1. *Log on to the Auto Scaling console*.

2. **Click the Monitoring Tasks tab and click Create in the upper-right corner. On the Create Monitoring Task page that you are redirected to, set the parameters.**

   The *Parameters and descriptions* table lists the parameters and descriptions.

   Table 4-8: Parameters and descriptions

Parameter	Description
Region	Required. Select the region where you want to create the monitoring task.
Department	Required. Select the department that the monitoring task belongs to.
Project	Required. Select the project that the monitoring task belongs to.
Task Name	Required. The name must be from 2 to 40 characters in length and can contain underscores (_), hyphens (-), or periods (.). It must start with a number, letter (case-insensitive), or Chinese character.
Description	Enter a description for the monitoring task. The description must be at least two characters in length.
Resource Monitored	Required. Select the scaling group that you want to monitor.
Reference Period	The reference period is measured in minutes. During a reference period, data is collected, summarized, and analyzed. The shorter the reference period, the more frequently the alert rules are triggered. Valid values: · 1 · 2 · 5 · 15
Metric	Required. Select the metric that you want to monitor. Options: · CPU · Internal Network Inbound Traffic · Internal Network Outbound Traffic · System Load Average · Memory

Parameter	Description
Condition	Required. The rule that determines whether to trigger alerts. Set this parameter to Average, Maximum, or Minimum, and then specify a threshold. In the following examples, alerts are triggered when the CPU utilization exceeds 80%: · Average: Alerts are triggered when the average CPU utilization of the ECS instances in the scaling group exceeds 80%. · Maximum: Alerts are triggered when the highest CPU utilization among the ECS instances in the scaling group exceeds 80%. · Minimum: Alerts will be triggered when the lowest CPU utilization among the ECS instances in the scaling group exceeds 80%.
Trigger After	This parameter specifies the number of times that the threshold can be reached within a reference period before alerts are triggered. Valid values: · 1 · 2 · 3 · 5
Triggered Rules	Required. Alert rules determine whether to perform scale-up or scale-down activities. You can add up to five alert rules to each monitoring task. Note: We recommend that you select scaling rules from the monitored scaling group.

3. Click OK to create the monitoring task.

4.9.3 View monitoring task details

This topic describes how to view details of a monitoring task in the Auto Scaling console.

Procedure

1. *Log on to the Auto Scaling console*.

2. Click the Monitoring Tasks tab and find the target monitoring task. In the Actions column, click the icon.
3. Select View Details. A dialog box appears displaying the details of the monitoring task.

   The details include basic information about the monitoring task, alert rules, and monitoring information.

4.9.4 Stop a monitoring task

This task describes how to stop a monitoring task in the Auto Scaling console.

Procedure

1. *Log on to the Auto Scaling console*.
2. Click the Monitoring Tasks tab and find the target monitoring task. In the Actions column, click the icon and select Stop.
3. In the dialog box that appears, click OK.

4.9.5 Start a monitoring task

This topic describes how to start a monitoring task in the Auto Scaling console.

Prerequisites

The target monitoring task is in the Stopped state.

Procedure

1. *Log on to the Auto Scaling console*.
2. Click the Monitoring Tasks tab and find the target monitoring task. In the Actions column, click the icon and select Start.
3. In the dialog box that appears, click OK.

4.9.6 Change monitoring task information

This topic describes how to change the basic information about a monitoring task.

Procedure

1. *Log on to the Auto Scaling console*.
2. Click the Monitoring Tasks tab and find the target monitoring task. In the Actions column, click the icon and select Change Basic Information.

3. **In the Change Basic Information dialog box that appears, you can change the following parameters:**

   Table 4-9: Parameters and descriptions

Parameter	Description
Monitoring Task Name	The name can be from 2 to 40 characters in length and can contain underscores (_), hyphens (-), or periods (.). It must start with a number, letter (case-insensitive), or Chinese character.
Description	Enter a description for the monitoring task. The description must be at least two characters in length.
Reference Period	The reference period is measured in minutes. The reference period indicates the time period during which data is collected, summarized, and analyzed. The shorter the reference period, the more frequently the alert rules are triggered. Valid values: · 1 · 2 · 5 · 15
Metric	Select the metric that you want to monitor. Options: · CPU · Internal Network Inbound Traffic · Internal Network Outbound Traffic · System Load Average · Memory
Condition	The rule that determines whether to trigger alerts. Set this parameter to Average, Maximum, or Minimum, and specify a value as the threshold.
Trigger After	This parameter specifies the number of times that the threshold can be reached within a reference period before alerts are triggered. Valid values: · 1 · 2 · 3 · 5

4. **Click OK.**

4.9.7 Change alert rules

This topic describes how to change alert rules for a monitoring task in the Auto Scaling console.

Procedure

1. *Log on to the Auto Scaling console*.
2. **Click the Monitoring Tasks tab and find the target monitoring task. In the Actions column, click the icon and select Change Triggered Rule.**
3. **In the Change Triggered Rule dialog box that appears, select an alert rule from Available Alert Rules and add it to Selected.**

   Note:
   You can add up to five alert rules to each monitoring task. We recommend that you select scaling rules from the monitored scaling group.
4. **Click OK.**

4.9.8 Delete a monitoring task

This topic describes how to delete monitoring tasks that you no longer need in the Auto Scaling console.

Procedure

1. *Log on to the Auto Scaling console*.
2. **Click the Monitoring Tasks tab and find the target monitoring task. In the** `Actions` **column, click the icon and select Delete.**
3. **Click OK.**

5 Object Storage Service (OSS)

5.1 What is OSS?

Alibaba Cloud Object Storage Service (OSS) is a massive, secure, low-cost, and highly reliable cloud storage service provided by Alibaba Cloud.

It can be considered as an out-of-the-box storage solution with unlimited storage capacity. Compared with the user-created server storage, OSS has many outstandin g advantages in reliability, security, cost, and data processing capabilities. Using OSS, you can store and retrieve a variety of unstructured data files, such as text files , images, audios, and videos, over the network at any time.

OSS uploads data files as objects to buckets. OSS is an object storage service that uses a key-value pair format. You can retrieve object content based on unique object names (keys).

On OSS, you can:

- **Create a bucket and upload objects to the bucket.**
- **Obtain an object URL from OSS to share or download an object.**
- **Complete the ACL settings of a bucket or object by modifying its properties or metadata.**
- **Perform basic and advanced OSS tasks through the OSS console.**
- **Perform basic and advanced OSS tasks using the Alibaba Cloud SDKs or directly calling the RESTful APIs in your application.**

5.2 Instructions

Before you use OSS, you need to understand the following content:

- **To allow other users to use all or part of OSS functions, you need to create RAM users and grant permissions to the users by assigning RAM policies to them. For more information, see content related to RAM users and RAM policies in RAM User Guide.**

- **Before you use OSS, you also need to understand the following service limits.**

Item	Description
Bucket	- You can create a maximum of 10 buckets. - After a bucket is created, its name and region cannot be modified.
Object upload	- Objects uploaded through the console, simple upload, form upload, or append upload cannot exceed 5 GB. To upload an object greater than 5 GB, you must use multipart upload. The size of an object that you want to upload in the multipart upload mode cannot exceed 48.8 TB. - You can upload an object with the same name as an existing object, but the existing object is overwritten.
Object deletion	- Deleted objects cannot be restored. - You can delete up to 50 objects at a time in the OSS console. To delete more than 50 objects at a time, you must call an API operation or use an SDK.
Lifecycle	You can configure up to 1,000 lifecycle rules for each bucket.
Image processing (IMG)	- Source image: ■ Only JPG, PNG, BMP, GIF, WebP, and TIFF objects are supported. ■ The object size cannot exceed 20 MB. ■ If you use image rotation, the width or height of the image cannot exceed 4,096 pixels. - Thumbnail: ■ The product dimensions cannot exceed 4,096 × 4,096 pixels. ■ The length of each side cannot exceed 4,096 pixels.

5.3 Quick start

5.3.1 Log on to the OSS console

This topic describes how to log on to the OSS console.

Prerequisites

- **Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://*IP address or domain name of the Apsara Stack console*/manage.**
- **We recommend that you use the Chrome browser.**

Procedure

1. **Open your browser.**
2. **In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name of the Apsara Stack console*/manage, and then press Enter.**
3. **Enter the correct username and password.**
   - **The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.**
   - **You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).**
4. **Click LOGIN to go to the Dashboard page.**
5. **In the left-side navigation pane, choose Compute, Storage & Networking > Object Storage Service.**

5.3.2 Create buckets

Objects uploaded to OSS are stored in buckets. Before you upload an object to OSS, you need to create a bucket.

Context

Properties of a bucket include the region, ACL settings, and other metadata.

Procedure

1. *Log on to the OSS console*.
2. **Click Create Bucket. In the Add Bucket dialog box that appears, set parameters as required.**

   Table 5-1: Parameters for creating a bucket **lists the parameters for creating a bucket.**

   Table 5-1: Parameters for creating a bucket

Parameter	Configuration method
Department	Select a department from the drop-down list.
Project	Select a project from the drop-down list.
Region	Select the data center where the bucket is deployed from the drop-down list. **Note:** · After a bucket is created, the region cannot be changed. · If you want to access OSS from your ECS instance through the internal network, select the same region where your ECS instance is deployed.

Parameter	Configuration method
Permissions	Set the ACL for the bucket. The following options are available: · Private: Only the owner or authorized user can read from and write to objects. · Public (Read-Only): Only the owner or authorized user can read from and write to objects. Other users (including anonymous users) can only read from files. · Public: Everyone can read from or write to objects in the bucket. Fees incurred by such operations are paid by the owner of the bucket. Configure this option only when necessary. Note: After a bucket is created, you can modify its ACL. For more information, see *Change ACL settings*.
Bucket Name	Enter the name of the bucket. Note: · The bucket name must comply with the naming conventions. · The bucket name must be globally unique in Apsara Stack OSS. · The bucket name cannot be changed after the bucket is created.
Bucket Capacity	Configure the capacity of the bucket.
Instances	Enter the number of buckets that you apply for. You can create a maximum of 10 buckets at a time.

3. Click Create. The bucket is created.

5.3.3 Upload objects

After you create a bucket, you can upload objects to it.

Context

You can upload an object of any format to a bucket. You can use the OSS console to upload an object no larger than 5 GB to a bucket. To upload an object larger than 5 GB, use an SDK or call an API operation.

Procedure

1. *Log on to the OSS console*.
2. **On the OSS homepage, click the name of the bucket to which an object is to be uploaded. The Bucket Information tab appears.**
3. **Click the Object Management tab. The list of objects appears.**
4. **Click Upload File.**
5. **In the dialog box that appears, select the object to be uploaded and click Open.**
6. **After the object is uploaded, refresh the Object Management tab to view the uploaded object.**

   You can view the upload progress and result on the Task Management tab.

5.3.4 Obtain object URLs

You can obtain the URL of an object uploaded to a bucket. This URL can be used to share or download the object.

Prerequisites

Before you obtain an object URL, you need to create a bucket and upload an object to it.

Procedure

1. *Log on to the OSS console*.
2. **Click the name of a bucket. The Bucket Information tab appears.**
3. **Click the Object Management tab. The list of objects appears.**
4. **Click the**  **icon in the Actions column corresponding to an object and choose Get URL from the shortcut menu. . The Get Object URL dialog box appears.**

 **Note:**

To obtain the URL of a bucket that has the ACL of Private, you need to configure a validity period for the URL. Click Get URL to obtain the object URL. The validity period of a signed URL is calculated based on NTP. You can share the object URL with other users so that they can use the URL to access the object within the validity period. If the bucket ACL is Private, the obtained URL is a signed URL.

5. **Copy the object URL and send it to other users so that they can view or download the object.**

5.4 Buckets

5.4.1 View a bucket

You can view the details of created buckets in the OSS console.

Prerequisites

Before you view a bucket, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Procedure

1. ***Log on to the OSS console*.**
2. **Click the name of the target bucket, or click the  icon in the Actions column and then click Details.**

   On the bucket details page that appears, click the Bucket Information tab. On the Bucket Information tab page, view bucket details such as Service IP Address and Creation Time.

5.4.2 Delete buckets

You can delete buckets in the OSS console.

Prerequisites

Before you delete a bucket, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

 **Note:**

To delete a bucket, make sure that all objects in it are deleted, including parts generated from incomplete multipart upload. Otherwise, the bucket cannot be deleted.

Procedure

1. *Log on to the OSS console*.
2. Click the  icon in the Actions column corresponding to the bucket to be deleted and choose Delete from the shortcut menu.
3. In the Delete Bucket message that appears, click OK.

 **Note:**
To delete multiple buckets at a time, select these buckets and click Delete. In the Delete Bucket message that appears, click OK.

5.4.3 Change the capacity

During actual usage, you may need to scale up or down the capacity of a bucket. You can change the capacity of a bucket in the OSS console.

Prerequisites

Before you change the capacity of a bucket, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Procedure

1. *Log on to the OSS console*.
2. Click the  icon in the Actions column of the target bucket and click Change Capacity.
3. In the Change Capacity dialog box that appears, change the capacity of the bucket. Click OK.

5.4.4 Change the ownership

You can change the department or project to which a bucket belongs in the OSS console.

Prerequisites

Before you change the department or project to which a bucket belongs, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Procedure

1. *Log on to the OSS console*.
2. Click the  icon in the Actions column of the target bucket and click Change Ownership.
3. In the Change Ownership dialog box that appears, change the department or project to which the bucket belongs. Click OK.

5.4.5 Change ACL settings

You can change Access Control List (ACL) settings of a bucket in the OSS console to control access to the bucket.

Prerequisites

Before you change the ACL settings of a bucket, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Context

OSS provides the ACL function to control bucket access permissions. After a bucket is created, the ACL of the bucket is set to private by default. You can change ACL settings after creating a bucket.

The OSS ACL function provides bucket-level access control. Three ACL settings are available for a bucket:

- Private: Only the bucket owner and authorized users can perform read and write operations on objects in the bucket. Other users cannot access objects in the bucket without authorization.
- Public (Read-Only): Only the bucket owner and authorized users can perform write operations on objects in the bucket. Other users (including anonymous users) can perform read operations on objects in the bucket.
- Public: All users (including anonymous users) can perform read and write operations on objects in the bucket. Fees incurred by these operations are paid by the bucket owner. Configure this permission type only when necessary.

Procedure

1. *Log on to the OSS console*.
2. Click the name of the target bucket, or click the  icon in the Actions column and then click Details.
3. On the bucket details page that appears, click the Bucket Properties tab. On the Bucket Properties tab page, click Read/Write Permissions.
4. Select an option for Read/Write Permissions.
5. Click Set to save your modifications.

5.4.6 Configure static website hosting

You can configure static website hosting in the OSS console so that users can access the static website from the bucket endpoint.

Prerequisites

Before you configure static website hosting for a bucket, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Context

If the default page is blank, static website hosting is disabled.

The default homepage is displayed if you directly access the root domain name of the static website or any URL ending with a forward slash (/) under this domain name.

Procedure

1. *Log on to the OSS console*.
2. Click the name of the target bucket, or click the  icon in the Actions column and then click Details.

3. On the bucket details page that appears, click the Bucket Properties tab. On the Bucket Properties tab page, click Website Settings.

   Configure the following parameters:

   - Default Homepage indicates the index page (equivalent to index.html of a website). You must enter the name of an HTML object that is stored in the bucket.
   - Default 404 Page indicates the default 404 page that is displayed when you access an incorrect path. You must enter the name of an HTML object that is stored in the bucket. If this field is left empty, the default 404 page is disabled.

4. Click Set to save your settings.

5.4.7 Enable logging

You can enable or disable logging for a bucket in the OSS console.

Prerequisites

Before you enable or disable logging for a bucket, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Context

You can store access logs in the current bucket or a new bucket.

Procedure

1. *Log on to the OSS console*.
2. Click the name of a bucket, or click the  icon in the Actions column corresponding to the bucket and choose View Details from the shortcut menu.
3. On the Bucket Information tab that appears, click the Bucket Properties tab. On the Bucket Properties tab that appears, click Logging Settings.

   Configure the following parameters:

   - Select the name of the bucket where access logs are to be stored from the Log Store drop-down list. The selected bucket must be owned by you and in the same region as the bucket for which logging is enabled. You can select Do Not Save to disable logging.
   - Enter a prefix in the Log Prefix field. This parameter corresponds to `<TargetPrefix>` in the following naming conventions. Access logs are

stored in the root directory. You can also add a folder path in front of *<TargetPrefix>*, such as *log/<TargetPrefix>*. Access logs are stored in the *log/* directory.

The naming conventions for objects that store access log entries are as follows:

<TargetPrefix><SourceBucket>YYYY-MM-DD-HH-MM-SS-<UniqueString>

- *<TargetPrefix>*: specifies the specified log prefix.
- *<SourceBucket>*: specifies the name of the source bucket.
- *YYYY-MM-DD-HH-MM-SS*: specifies the time in Coordinated Universal Time (UTC +8) when the access log is created. YYYY specifies a 4-digit year. MM specifies a 2-digit month. DD specifies a 2-digit day. HH specifies a 2-digit hour. MM specifies a 2-digit minute. SS specifies a 2-digit second.
- *<UniqueString>*: specifies a string generated by OSS.

For example, the name of an object that stores OSS access log entries is MyLog-OSS-example2015-09-10-04-00-00-0000.

MyLog- indicates the specified log prefix. oss-example indicates the name of the source bucket. 2015-09-10-04-00-00 indicates the time in UTC+8 when the access logs are created. 0000 indicates a string generated by OSS.

4. Click Set to save your settings.

5.4.8 Configure hotlink protection

You can configure hotlink protection for a bucket in the OSS console to prevent other domain names from accessing the data in your bucket.

Prerequisites

Before you configure hotlink protection for a bucket, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Context

OSS provides hotlink protection to prevent other domain names from accessing your data in OSS. You can configure the Referer field in the HTTP header to implement hotlink protection. In the OSS console, you can configure a whitelist for the Referer field and configure whether to allow requests with an empty Referer field. For example, you can add http://www.aliyun.com to the Referer whitelist for a

bucket named oss-example. Then, requests with a Referer field value of http://www.aliyun.com can access the objects in the oss-example bucket.

Procedure

1. *Log on to the OSS console*.
2. Click the name of a bucket, or click the  icon in the Actions column corresponding to the bucket and choose View Details from the shortcut menu.
3. On the Bucket Information tab that appears, click the Bucket Properties tab. On the Bucket Properties tab that appears, click Anti-leech Settings.
4. Add whitelist URLs to the Referer field.
5. Configure whether to allow requests with an empty Referer field.

   Select Allow Empty Referer Field if you do not need to limit access requests.
6. Click Submit to save your settings.

5.4.9 Configure CORS

You can configure cross-origin resource sharing (CORS) in the OSS console to enable cross-origin access.

Prerequisites

Before you configure CORS for a bucket, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Context

OSS provides CORS in HTML5 to enable cross-origin access. When OSS receives a cross-origin request (or OPTIONS request), OSS reads the CORS rules of the destination bucket and checks its ACL setting. OSS matches the rules one by one. When OSS finds the first match, it returns a corresponding header. If no matches are found, OSS does not attach any CORS-related headers.

Procedure

1. *Log on to the OSS console*.
2. Click the name of the bucket to be accessed, or click the  icon in the Actions column corresponding to the bucket and choose View Details from the shortcut menu.

3. **On the Bucket Information tab that appears, click the Bucket Properties tab. On the Bucket Properties tab that appears, click CORS Rules.**
4. **Click Create Rule.**
5. **In the Add CORS Settings dialog box that appears, set parameters as required.**

   Table 5-2: Parameters for CORS settings **lists parameters for CORS settings.**

   Table 5-2: Parameters for CORS settings

Parameter	Description
Source	Specifies the sources you want to allow cross-origin requests from. You can configure multiple matching rules separated by carriage returns. Each matching rule can contain up to one asterisk (*) as the wildcard.
Method	Specifies the allowed CORS request methods.
Allowed Header	Specifies the allowed headers in a cross-origin request. You can configure multiple matching rules separated by carriage returns. Each matching rule can contain up to one asterisk (*) as the wildcard.
Expose Header	Specifies the response headers that allow access from applications.
Cache Time	Specifies how long the browser can cache the response to a preflight (OPTIONS) request to a specific resource.

   Note:

   You can configure up to 10 rules for each bucket.

6. **Click OK to save your settings. You can also edit or delete existing rules.**

5.4.10 Manage lifecycle rules

You can define and manage lifecycle rules for a bucket in the OSS console.

Prerequisites

Before you manage lifecycle rules of a bucket, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Context

You can define a rule for a full set or a subset (by specifying the prefix keyword) of objects in a bucket. A rule is automatically applied to all objects that match the rule. You can use lifecycle management to perform operations, such as batch object management and automatic part deletion.

- **If an object matches a rule, data of the object is cleared within two days from the effective date.**
- **Data that is deleted in batches based on a lifecycle rule cannot be restored. Configure a rule only when necessary.**

Procedure

1. ***Log on to the OSS console*.**
2. **Click the name of the target bucket, or click the**  **icon in the Actions column and then click Details.**
3. **On the bucket details page that appears, click the Bucket Properties tab. On the Bucket Properties tab page, click LifeCycle Settings.**
4. **Click Add Rules.**
5. **In the Add LifeCycle Rule dialog box, configure required parameters.**

   ***Table 5-3: Parameters for lifecycle settings* lists the parameters for lifecycle settings.**

   Table 5-3: Parameters for lifecycle settings

Parameter	Description
Status	Select the status of this rule. You can select Enable or Disable.
Policy	Select an object matching policy, including Configure for Entire Bucket and Configure by Prefix.

Parameter	Description
Prefix	If image objects with the prefix img/ are stored in the bucket, you can enter `img/` in this field to manage the lifecycle of these objects.
Expired	Configure an expiration date or days to expiration for objects. · Set by Date: indicates the expiration date of objects. All objects that are created before this date are deleted. Perform this operation only when necessary. · Set by Number of Days: indicates days to expiration, that is, a lifecycle of objects in days. When the number of days from the last modification of an object exceeds the specified number of days, the object is deleted based on the rule. If this parameter is set to 30 days, objects last modified on January 1, 2016 are deleted on January 31, 2016.

6. Click Confirm to save your settings. After the rule is saved, you can view it on the LifeCycle Settings tab page. You can click Edit or Delete in the Actions column to edit or delete the rule.

5.4.11 Configure cross-cloud replication

You can synchronize data in a bucket from the OSS console to another cloud.

Prerequisites

Before you configure cross-cloud synchronization, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region and there is data uploaded to the bucket.

Context

You can specify a prefix to allow only objects with this prefix to be replicated to another cloud. You can configure a data synchronization policy by specifying a synchronization type and historical data synchronization.

Procedure

1. *Log on to the OSS console*.
2. Click the name of a bucket, or click the  icon in the Actions column and choose View Details from the shortcut menu.
3. On the Bucket Information tab that appears, click the Bucket Properties tab. On the Bucket Properties tab that appears, click Cross-Cloud Replication.
4. On the Cross-Cloud Replication tab that appears, click Enable Data Synchronization. The Enable Data Synchronization dialog box appears.
5. Configure parameters for cross-cloud synchronization.

   Table 5-4: Parameters for cross-cloud synchronization lists the parameters for cross-cloud synchronization.

   Table 5-4: Parameters for cross-cloud synchronization

Parameter	Configuration method
Destination Cloud	Select the destination cloud for data synchronization.
Destination Cloud Address	Enter the address of the destination cloud.
Destination Bucket	Enter the name of the bucket in the destination cloud.  **Note:** The name of the source bucket is added by the system. Cross-cloud data synchronization replicates data from the source bucket to the bucket in the destination cloud. Therefore, you need to specify the bucket in the destination cloud. The destination bucket must have the same name as the source bucket.
Source Object	Select the object in the source bucket. You can select Synchronize All. You can also select Synchronize with Prefix, click Add, and enter a prefix to synchronize objects with this prefix.

Parameter	Configuration method
Synchronization Policy	Select a synchronization policy. You can select Write Synchronization (Add/Delete) or Add/Delete/Change Synchronization.
Synchronize Historical Data	Select whether to synchronize historical data. You can set whether to synchronize data at a point of time before data is modified.

6. Click OK to save your settings. After your settings are saved, you can view this rule on the Cross-Cloud Replication tab. You can click Edit or Delete in the Actions column to edit or delete the rule.

5.5 Object

5.5.1 Search for objects

You can search buckets or folders for objects with a specific name prefix in the OSS console.

Prerequisites

Before you search for an object, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region and at least one object in the bucket.

Context

When you search for an object based on a prefix, the search string is case-sensitive and cannot contain a forward slash (/). The search range is limited to the root directory of the current bucket or the objects in the current folder (excluding subfolders and the objects in them).

Procedure

1. *Log on to the OSS console*.
2. Click the name of the target bucket, or click the  icon in the Actions column and then click Details.

3. On the bucket details page that appears, click the Object Management tab.
4. On the Object Management tab page, enter a prefix in the search box, and press Enter or click Search.

   To search a folder, open the folder and enter a prefix in the search box. The system lists the names of objects and folders matching the prefix in the root directory of the folder.

5.5.2 Delete objects

You can delete uploaded objects in the OSS console.

Prerequisites

Before you delete objects, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region and at least one object in the bucket.

Context

You can delete one or more objects at a time. A maximum of 1,000 objects can be deleted at a time. You can use an SDK or call an API operation to delete a specific object or more than 1,000 objects.

Notice:
Deleted objects cannot be restored. Exercise caution when you delete objects.

Procedure

1. *Log on to the OSS console*.
2. Click the name of the bucket in which the object is to be deleted, or click the 

   icon in the Actions column corresponding to the bucket and choose View Details from the shortcut menu.
3. On the Bucket Information tab that appears, click the Object Management tab.
4. On the Object Management tab that appears, click the  icon in the Actions

   column corresponding to the object and choose Delete from the shortcut menu.

   Notice:
   A folder may fail to be deleted if it contains an excessive number of objects.

5. In the Delete Object message that appears, click OK.

5.5.3 Configure ACL of an object

You can set the ACL for an object in the OSS console to control access to the object.

Prerequisites

Before you configure ACL of an object, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region and at least one object in the bucket.

Procedure

1. *Log on to the OSS console*.
2. Click the name of the target bucket, or click the  icon in the Actions column and then click Details.
3. On the bucket details page that appears, click the Object Management tab. The Object Management tab page is displayed.
4. Click the  icon in the Actions column corresponding to the target object and click Set File ACL.
5. In the Set File ACL dialog box that appears, select an option from the Read/Write Permissions drop-down list.
6. Click OK.

5.5.4 Create folders

You can create a folder in a bucket in the OSS console.

Prerequisites

Before you create a folder, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Context

OSS does not use folders. All elements are stored as objects. In the OSS console, a folder is an object with the size of 0 and has a name that ends with a forward slash (/). A folder is used to sort objects of the same type and process them simultaneously. The OSS console displays objects that end with a forward slash (/) as folders by default. These objects can be uploaded and downloaded normally. You can use OSS folders in the OSS console in the way you use folders in Windows.

Note:
The OSS console displays any objects that end with a forward slash (/) as a folder, regardless of whether these objects contain data. You can download such objects only by calling an API operation or using an SDK.

Procedure

1. *Log on to the OSS console*.
2. Click the name of a bucket, or click the  icon in the Actions column corresponding to the bucket and choose View Details from the shortcut menu.
3. On the Bucket Information tab that appears, click the Object Management tab.
4. On the Object Management tab that appears, click Create Folder.
5. In the Create Folder dialog box that appears, enter a folder name in the Folder Name field.
6. Click OK.

5.6 Image service

5.6.1 Create styles

You can create an image style in the OSS console to define a processing rule for uploaded image objects.

Prerequisites

Before you create an image style, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Procedure

1. *Log on to the OSS console*.
2. Click the name of a specified bucket, or click the  icon in the Actions column corresponding to the bucket and choose View Details from the shortcut menu.
3. On the Bucket Information tab that appears, click the Image Processing tab.

4. **On the Image Processing tab, click Create Style.**

   Table 5-5: Parameters for creating a style **lists the parameters for creating a style.**

   Table 5-5: Parameters for creating a style

Parameter	Configuration method
Rule Name	**Enter a style name, which must comply with the naming conventions.**
Change Type	**You can select Basic to edit the image style based on the graphical user interface (GUI). You can also select Advanced to edit the image style using an SDK or setting parameters.**
Preview	**Select an image preview method.**
Resize	**Select a resizing method for the thumbnail.** **There are two resizing methods:** · **Resize with Fixed Height and Width** - **Scale by shorter edge, center and crop.** - **Scale by longer edge, fill empty space.** · **Maintain Aspect Ratio** **Note:** **A longer edge refers to the side that is larger than the required size ratio. A shorter edge refers to the side that is shorter than the required size ratio. For example, the size of a source image is 400 × 200 and is scaled to 800 × 100. The ratios are calculated as follows: 400/800 = 0.5. 200/100 = 2. 200 is the longer edge and 400 is the shorter edge because 0.5 is smaller than 2.**
Thumbnail Width	**Configure the thumbnail dimension in px.**

Parameter	Configuration method
Limit	Configure whether to disable enlarging .
Auto Rotation	Configure Auto Rotation for the image, including Not Rotation and Rotation.
Image Processing	Specify whether the image requires special processing. You can select Sharpening.
Quality	Configure image quality, including Relative, Absolute, and No Compression.
Save as Format	Configure the format in which the image is saved, including Original Format, JPG, PNG, WebP, and BMP.
Add Watermark	Configure the watermark mode of the image, including No Watermark, Text Watermark, and Image Watermark.

5. After editing the image style, click Submit to save your settings.
6. After the style is submitted, click Export Style to download the style to your local device.

5.6.2 Enable source image protection

You can enable source image protection in the OSS console to prevent unauthorized use of images.

Prerequisites

Before you enable source image protection, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Context

To prevent unauthorized use of images in business systems, you need to prevent exposure of image URLs. Thus, unauthorized users can only obtain thumbnails or watermarked images. For this purpose, you can enable source image protection . After source image protection is enabled, source images are accessible only through URLs carrying stylenames or signature-based accesses. You are not

allowed to access source images in OSS or specify image parameters to modify image styles.

Procedure

1. *Log on to the OSS console*.
2. Click the name of the target bucket, or click the  icon in the Actions column and then click Details.
3. On the bucket details page that appears, click the Image Processing tab. On the Image Processing tab page, click Service Management.
4. Click Edit. Set Source Image Protection.

   If you select Enable, you must also set File Extensions for Source Image Protection to restrict access to source images with one or more extensions.

   Source image protection is designed to protect image objects. You must specify the extensions of image objects to be protected. For example, if you enable source image protection for `.jpg` objects, you can still directly access the source images of `.png` objects.
5. In Style Access Method, set Delimiter (Default: @!).
6. Click Save to save your settings.

5.7 Create single tunnels

You can create single tunnels between OSS and a VPC so that you can access OSS resources from the VPC.

Prerequisites

Before you can create single tunnels, you need to create a VPC and a VSwitch.

Procedure

1. *Log on to the OSS console*.
2. Click the OSS Access Control for VPC tab.
3. On the OSS Access Control for VPC tab page, click Create Single Tunnel.

4. **In the Create Single Tunnel dialog box, set required parameters.**

   Table 5-6: Parameter for creating single tunnels **lists the parameter configurations for creating single tunnels.**

   Table 5-6: Parameter for creating single tunnels

Parameter	Description
Region	**Select a region.**
Department	**Select a department or all department s.**
Description	**Enter a description for the single tunnel.**
VPC	**Select a VPC. For more information about how to create a VPC, see "Create a VPC and a VSwitch" in** *VPC User Guide***.**
VSwitch	**Select a VSwitch. For more information about how to create a VSwitch, see " Create a VSwitch" in** *VPC User Guide***.**

5. **Click Confirm.**

6 Table Store

6.1 What is Table Store?

Table Store is a NoSQL database service independently developed by Alibaba Cloud. Table Store is a proprietary software program that is certified by the relevant authority in China. Table Store is built on the Apsara system of Alibaba Cloud, and can store large amounts of structured data and allow real-time access to these data.

Table Store provides the following features:

- **Table Store offers schema-free data structure storage. You do not need to define attribute columns before you use them. You do not require table-level changes to add or reduce attribute columns. You can enable time to live (TTL) on a table to delete expired data from the table.**
- **Adopts the triplicate technology to keep three copies of data on three servers across three different racks. Each cluster supports either pure SSD instances or mixed storage instances (SSD and SATA) to meet different budget and performance requirements.**
- **Adopts a fully redundant architecture that prevents single point of failures (SPOFs). With support for online smooth upgrades, hot cluster upgrades, and automatic data migration, you can dynamically add or remove nodes without service interruptions for maintenance. The concurrent read/write throughput and storage capacity can be linearly scaled. Each cluster can have no less than 500 hosts.**
- **Supports highly concurrent read/write operations. Concurrent read/write capabilities can be scaled out with the increase of hosts. The read/write performance is indirectly related to the amount of data in a single table.**
- **Supports identity authentication and multi-tenancy. Comprehensive access control and isolation mechanisms are provided to safeguard your data. VPC and access over HTTPS are supported. Provides multiple authentication and authorization mechanisms so that you can define access permissions for individual tables and operations.**

6.2 Limits

Before using Table Store, you need to take note of the following precautions and limits.

The following table describes the limits for Table Store. Some of the limit ranges indicate the maximum allowable values instead of the suggested values. For better performance, set the table scheme and data size in a single row properly based on actual conditions, and adjust the following configurations as needed.

Item	Limit	Description
The number of instances under an Apsara Stack tenant account	1,024	To raise the limit, contact the technical support personnel.
The number of tables in an instance	1,024	To raise the limit, contact the technical support personnel.
Instance name length	3 to 16 Bytes	The instance name can contain uppercase and lowercase letters, digits, and hyphens (-). It must start with a letter and cannot end with a hyphen (-).
Table name length	1 to 255 Bytes	The table name can contain uppercase and lowercase letters, digits, and underscores (_). It must start with a letter or underscore (_).
Column name length	1 to 255 Bytes	The column name can contain uppercase and lowercase letters, digits, and underscores (_). It must start with a letter or underscore (_).
The number of columns in a primary key	1 to 4	A primary key can contain one to four columns.
The size of the value in a string type primary key column	1 KB	The size of the value in a string type primary key column cannot exceed 1 KB.
The size of the value in a string type attribute column	2 MB	The size of the value in a string type attribute column cannot exceed 2 MB.
The size of the value in a binary type primary key column	1 KB	The size of the value in a binary type primary key column cannot exceed 1 KB.

Item	Limit	Description
The size of the value in a binary type attribute column	2 MB	The size of the value in a binary type attribute column cannot exceed 2 MB.
The number of attribute columns in a single row	Unlimited	A single row can contain an unlimited number of attribute columns.
The number of attribute columns written by one request	1,024	During a PutRow, UpdateRow, or BatchWriteRow operation, the number of attribute columns written in a row cannot exceed 1,024.
The data size of a row	Unlimited	The total size of all column names and column values for a row is unlimited.

6.3 Quick start

6.3.1 Log on to the Table Store console

This topic describes how to log on to the Table Store console.

Prerequisites

- **Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://***IP address or domain name of the Apsara Stack console*/**manage.**
- **We recommend that you use the Chrome browser.**

Procedure

1. **Open your browser.**
2. **In the address bar, enter the access address of the Apsara Stack console in the format of http://***IP address or domain name of the Apsara Stack console*/**manage, and then press Enter.**

3. Enter the correct username and password.
   - The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click LOGIN to go to the Dashboard page.
5. In the left-side navigation pane, choose Compute, Storage & Networking > Table Store.

6.3.2 Create an instance

An instance is a logical entity in Table Store and is used to manage tables. An instance is the basic unit in the resource management system of Table Store. Table Store implements application access control and resource measurement at the instance level.

Procedure

1. Log on to the *Table Store console*.
2. On the Table Store page, click Create Instance.

   Note:
   You can create different instances to manage the tables for different business, or create different instances for development, tests, and production environments of the same business. Table Store allows you to create up to 1,024 instances under an Apsara Stack tenant account, and up to 1,024 tables in each instance by default.
3. Enter Instance Name, and select `Department`, `Project`, `Region`, and `Instance Specification`.

   Note:

- Table Store supports high-performance instances and capacity instances. The instance types vary with the cluster you deploy.
- Instance naming conventions: The name must be 3 to 16 characters in length and can only contain letters, digits, and hyphens (-). It must start with an uppercase or lowercase letter. It cannot start with `ali` or `ots`.

4. Click OK.

   The created instance is displayed on the Instances page.

6.3.3 Create a table

After creating an instance, you can create, update, and delete a table in the instance.

Procedure

1. Log on to the *Table Store console*.
2. Locate the instance you want to manage and click the instance name to go to the Instance Details tab.
3. Click the Data Tables tab.
4. On the Data Tables tab, click Create Data Table.

   Note:
   You can create a maximum of 64 data tables in an instance.

5. Specify data table information.

   The following table describes parameter configurations.

   Table 6-1: Data table parameters

Parameter	Description
Data Table Name	The data table name can contain uppercase and lowercase letters, numbers, and underscores (_). It must start with a letter or underscore (_). The data table names in an instance must be unique.

Parameter	Description
Reserved Read Throughput	The reserved read/write throughput can be set to 0. When the reserved read/write throughput is greater than 0, Table Store allocates and reserves corresponding resources for the table based on the configuration. The value ranges from 0 to 5,000 and must be an integer. Capacity-type instances do not support this parameter.
Reserved Write Throughput	
Data Life Cycle	The minimum data lifecycle is 86,400s (one day) or -1 (data never expires.).
Maximum Data Version	A non-zero value. It indicates the maximum number of data versions that can be stored in each attribute column of a data table. When the number of versions in an attribute column exceeds the parameter value, the earliest version is deleted. This operation is performed asynchronously.
Maximum Version Offset	The offset of the version of all written data columns from the data write time must be within the range of the valid version offset. Otherwise, data write may fail. The valid version range of an attribute column is calculated based on the formula: Valid version range = `[Data write time - Valid version offset, Data write time + Valid version offset)`.

Parameter	Description
Primary Key	A maximum of four primary keys can be set. The first primary key is the partition key by default. Click Add Primary Key to add a new primary key. The primary key type can be `Integer` or `String`. The primary key configuration and the key order cannot be modified after they are set. The primary key name can contain uppercase and lowercase letters, digits , and underscores (_). It must start with a letter or underscore (_).
Enable Stream	After Stream is enabled, the log is stored. The incremental data can be read through the channels provided by Stream during the specified storage life.

6. Click OK.

   The system automatically returns to the Data Tables tab and displays the table creation result. After the table is created, it is displayed on the Data Tables tab.

6.4 Manage instances

6.4.1 View an instance

In the Table Store console, you can view the region, creation time, and internal and external access URLs of an instance you have created.

Procedure

1. Log on to the *Table Store console*.

2. Click the icon in the Actions column corresponding to the instance you want to view, and choose View Details from the shortcut menu.

   The following information is displayed: the status, region, creation time, and internal and external access URLs of the instance, as well as whether a VPC is bound to the instance.

6.4.2 Release an instance

You can release a Table Store instance that you have created.

Prerequisites

Before releasing an instance, delete all tables from the instance. Otherwise, the instance cannot be released.

Procedure

1. Log on to the *Table Store console*.
2. Click the icon in the Actions column corresponding to the instance you want to release, and choose Release from the shortcut menu.
3. In the Delete message that appears, click OK.

6.5 Manage data tables

6.5.1 View details of data tables

You can view the basic information and actual usage of a table on the table management page.

Procedure

1. Log on to the Table Store console.
2. Locate the instance to be viewed, and click the instance name to go to the Instance Details page.
3. Click the Data Tables tab to go to the Data Tables page.
4. Click the icon in the Actions column corresponding to the instance you want to view, and choose View Details from the shortcut menu.

   On the page that appears, you can view the data table name, time to live (TTL), last modification time, primary keys (sorted in the sequence specified during table creation), and Stream information.

6.5.2 Update a table

You can adjust parameters for Table Store data tables, such as the time to live (TTL), max versions, and max version offset.

Procedure

1. **Log on to the *Table Store console*.**
2. **Locate the instance to be managed, and click the instance name to go to the Instance Details page.**
3. **Click the Data Tables tab.**
4. **Click the icon in the Actions column corresponding to the table you want to update, and choose Adjust Data Table Parameters from the shortcut menu.**
5. **Set the parameters to be updated, such as the time to live (TTL), max versions, and max version offset.**
6. **Click OK. On the Data Tables tab that appears, you can view the updated parameter values.**

6.5.3 Delete a table

You can delete a table you have created in the Table Store console.

Context

> **Notice:**
> **After a data table is deleted, the data in the table cannot be restored.**

Procedure

1. **Log on to the *Table Store console*.**
2. **Locate the instance you want to manage and click the instance name to go to the Instance Details tab.**
3. **Click the Data Tables tab.**
4. **Click the icon in the Actions column corresponding to the table you want to delete, and choose Release from the shortcut menu.**
5. **In the Confirm Deletion message, click Confirm.**

   After the deletion is confirmed, the table and the data in the table are deleted permanently.

6.6 Bind a VPC

Virtual Private Cloud (VPC) is an isolated network environment built in Apsara Stack. If you want to access a Table Store instance over a VPC, you must bind the VPC to the instance.

Prerequisites

- **You must create a VPC that is in the same region as the Table Store instance. For more information about how to create a VPC, see the Create a default VPC and VSwitch section in** *VPC User Guide* **.**
- **After the VPC is created, create an ECS instance in the VPC.**

Procedure

1. *Log on to the Table Store console*.
2. **Locate the instance to be bound to a VPC and click the instance name to go to the Instance Details page.**
3. **Click Bind VPC.**
4. **In the Bind VPC dialog box that appears, set VPC, VSwitch, and VPC Alias. Click OK.**

   Note:
   To allow a RAM user to log on to the Table Store console and manage VPCs, use an Apsara Stack tenant account to log on to the RAM console and grant the RAM user the relevant VPC permissions (AliyunVPCReadOnlyAccess). Otherwise, the RAM user cannot obtain VPC information in the Table Store console.
5. **After the instance is bound to the VPC, view the bound VPC information on the Instance Details page. Click the URL in the VPC ID column to go to the VPC**

Instance List page. The Table Store instances bound to the VPC and the VPC information are displayed.

You can use the VPC URL as the endpoint to access the Table Store instance from the ECS instance in that VPC.

If the VPC is no longer needed, click the icon in the Actions column corresponding to the VPC, and choose Unbind VPC from the shortcut menu to unbind the VPC from the instance.

After the VPC is unbound from the Table Store instance, the ECS instance in the VPC cannot access the Table Store instance through the preceding URL. To access the Table Store instance, you need to bind the VPC to the instance again.

7 Network Attached Storage (NAS)

7.1 What is NAS?

Alibaba Cloud Network Attached Storage (NAS) provides file storage services for compute nodes, such as ECS instances and Container Service nodes.

NAS supports multiple standard file access protocols. You can use a distributed file system that works seamlessly with existing applications while offering a variety of features. These features include unlimited capacity, scalable performance, unique namespace, shared access, high reliability, and high availability. Compared with traditional user-created data stores, NAS helps you reduce a large number of maintenance costs and mitigate data security risks. Furthermore, you can mount a NAS file system on multiple compute nodes at the same time. This helps you reduce a large number of costs in data transit and synchronization.

You can perform the following operations on a NAS file system:

- **Create NAS file systems and mount points.**
- **Create a permission group for a NAS file system and add rules to a permission group. This allows access to a file system from specified IP addresses or segments and allows you to grant different access permissions to different IP addresses or IP segments.**
- **Mount file systems on compute nodes, such as ECS instances and Container Service nodes by using the NFS or SMB protocol. Access a file system by using POSIX-based API operations.**
- **Manage file systems, mount points, and permission groups in the NAS console.**
- **Call API operations for NAS to manage resources.**

7.2 Instructions

Before you can use NAS, you need to understand the following content.

- **NAS supports the NFSv3 and NFSv4 protocols.**
- **NFSv4.0 does not support the following attributes: FATTR4_MIMETYPE, FATTR4_QUOTA_AVAIL_HARD, FATTR4_QUOTA_AVAIL_SOFT, FATTR4_QUO**

TA_USED, FATTR4_TIME_BACKUP, and FATTR4_TIME_CREATE. The client displays an NFS4ERR_ATTRNOTSUPP error.

- NFSv4.1 does not support the following attributes: FATTR4_DIR_NOTIF_DELAY , FATTR4_DIRENT_NOTIF_DELAY, FATTR4_DACL, FATTR4_SACL, FATTR4_CHA NGE_POLICY, FATTR4_FS_STATUS, FATTR4_LAYOUT_HINT, FATTR4_LAY OUT_TYPES, FATTR4_LAYOUT_ALIGNMENT, FATTR4_FS_LOCATIONS_ INFO, FATTR4_MDSTHRESHOLD, FATTR4_RETENTION_GET, FATTR4_RET ENTION_SET, FATTR4_RETENTEVT_GET, FATTR4_RETENTEVT_SET, FATTR4_RETENTION_HOLD, FATTR4_MODE_SET_MASKED, and FATTR4_FS_ CHARSET_CAP. The client displays an NFS4ERR_ATTRNOTSUPP error.
- NFSv4 does not support the following OPs: OP_DELEGPURGE, OP_DELEGRETURN , and NFS4_OP_OPENATTR. The client displays an NFS4ERR_NOTSUPP error.
- NFSv4 does not support Delegation.
- About UID and GID:
  - For NFSv3, if the file UID or GID exists in a Linux local account, the correspond ing username and group name are displayed based on the mapping between the local UID and GID. If the file UID or GID does not exist in the local account , the UID and GID are displayed.
  - For NFSv4, if the version of the local Linux kernel is earlier than 3.0, "nobody" is displayed as the UID and GID of all files. If the kernel version is later than 3. 0, the display rule is the same as that of NFSv3.

  Notice:
  If you use NFSv4 to mount a NAS instance and the Linux kernel version is earlier than 3.0, we recommend that you do not change the owner or group of local files or directories. Such changes can cause the UIDs and GIDs of the files or directories to become "nobody."
- A NAS instance can be mounted to up to 10,000 compute nodes for parallel access .

7.3 Quick start

7.3.1 Log on to the NAS console

This topic describes how to log on to the NAS console.

Prerequisites

- **Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://*IP address or domain name of the Apsara Stack console*/manage.**
- **We recommend that you use the Chrome browser.**

Procedure

1. **Open your browser.**
2. **In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name of the Apsara Stack console*/manage, and then press Enter.**
3. **Enter the correct username and password.**
   - **The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.**
   - **You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).**
4. **Click LOGIN to go to the Dashboard page.**
5. **In the left-side navigation pane, choose Compute, Storage & Networking > Network Attached Storage.**

7.3.2 Create a file system

A file system is the basic storage unit of NAS. Before using NAS, you must create a file system.

Context

When creating a file system, you must note the following precautions:

- **You can create a maximum of 1,000 file systems in each user account.**
- **The maximum capacity of a file system is 10 PB.**
- **If you need to increase the maximum capacity, we recommend that you contact Alibaba Cloud Technical Support.**

Procedure

1. *Log on to the NAS console*.
2. **In the NAS console, click Create File System.**
3. **In the Create File System dialog box, configure the required settings.**

   Table 7-1: Settings used to configure a file system **shows these settings and the description of each setting.**

   Table 7-1: Settings used to configure a file system

Category	Name	Description
Region	The region where the file system resides.	Select a region from the drop-down list.
Basic Settings	Department	Select a department from the drop-down list.
	Project	Select a project from the drop-down list.
	File System Name	Enter a name for the file system.
Storage Configuration	Storage Type	Select Capacity-type.
	Protocol Type	Select NFS or SMB.
	Quota (TB)	The size of the file system .

Note:

The quota of a NAS Capacity file system ranges from 1 to 10,240 TB. The quota must be an integer and cannot start with 0.

4. **Click OK to create a new file system.**

7.3.3 Create permission groups

NAS uses permission groups and permission group rules to manage NAS instance permissions. Before you can use a NAS instance, you must create permission groups and configure the required parameters.

Context

Each permission group in a NAS instance has an IP address whitelist. You can add rules to a permission group of a NAS instance to allow users from specified IP addresses or address segments to access the NAS instance with different permissions.

Procedure

1. *Log on to the NAS console*.
2. **On the File Storage NAS page, click the Permission Group tab.**
3. **On the Permission Group tab that appears, click Create Access Group.**

Note:
You can create up to 100 permission groups. If you need to raise the limit, contact the administrator.

4. **In the Create Permission Group dialog box, set the parameters.**

   Table 7-2: Parameters for creating a permission group **lists the parameters for creating a permission group.**

   Table 7-2: Parameters for creating a permission group

Parameter	Description
Region	Select a region from the drop-down list.
Department	Select a department from the drop-down list.
Project	Select a project from the drop-down list.

Parameter	Description
Permission Group Name	Enter the name of the permission group.
Network Type	Select VPC or Classic Network.

5. Click OK.

7.3.4 Create permission group rules

NAS uses permission groups and permission group rules to manage NAS instance permissions. Before you can use a NAS instance, you must create rules in its permission groups and configure the required parameters.

Context

Each permission group in a NAS instance has an IP address whitelist. You can add rules to a permission group of a NAS instance to allow users from specified IP addresses or address segments to access the NAS instance with different permissions.

> **Warning:**
> To ensure data security, we strongly recommend that you use caution when adding permission group rules and granting permissions to IP addresses.

Procedure

1. *Log on to the NAS console*.
2. On the File Storage NAS page, click the Permission Group tab. On the Permission Group tab that appears, click the name of a permission group to go to the Rules List page.
3. Click Create Rule.

> **Note:**
> You can create up to 1,000 permission group rules. If you want to raise the limit, contact the administrator.

4. **In the Add Rule dialog box, set the parameters.**

   Table 7-3: Parameters for creating a permission group rule **lists the parameters for creating a permission group rule.**

   Table 7-3: Parameters for creating a permission group rule

Parameter	Description
Authorized IP Address	The IP address or IP address segment of the object authorized by the rule. For a classic network, you can specify only one IP address.
Read/Write Permissions	Select Read-Only or Read and Write to allow the authorized object to perform read-only or read/write operations on the NAS instance.
User Permissions	Select Do Not Limit root User (no_squash), Limit root User (root_squash), or Limit All Users (all_squash) to specify whether to limit the access from the Linux system users of the authorized object to the NAS instance. Description: · Do Not Limit root User (no_squash) allows the root user to access the NAS instance. · Limit root User (root_squash) considers the root user as nobody. · Limit All Users (all_squash) considers all users including root as nobody.
Priority	The priority value ranges from 1 to 100. The value 1 indicates the highest priority. When an authorized object matches multiple rules, the rule with the highest priority takes effect.

5. **Click OK.**

7.3.5 Add mount points

After you create a NAS instance and its permission groups, you must add mount points to the NAS instance so that you can mount the NAS instance to compute nodes, such as ECS, E-HPC, or Container Service instances.

Context

A mount point is an access address of a NAS instance in a VPC or classic network. Each mount point corresponds to a domain name. NAS supports two types of mount points: VPC and classic network.

Procedure

1. *Log on to the NAS console*.
2. On the File Storage NAS page, click a NAS instance ID to go to the instance details page.
3. Click the Mount Point tab.
4. On the Mount Point tab that appears, click Add Mount Point.

   Note:
   You can create up to 100 mount points. If you need to raise the limit, contact the administrator.

5. In the Add Mount Point dialog box, configure the parameters.
   - If you set Mount Point Type to Classic Network, select a permission group to be bound to the mount point from the Permission Group drop-down list.
   - If you set Mount Point Type to VPC, set VPC and VSwitch. Then, select a permission group to be bound to the mount point from the Permission Group drop-down list.

   Note:
   - If you set Mount Point Type to VPC, ensure that the corresponding VPC and VSwitch have been created.
   - If you set Mount Point Type to Classic Network, the NAS instance can be accessed only by ECS instances under the same account as the mount point.
   - You can use a single mount point to mount a NAS instance to multiple compute nodes such as ECS, E-HPC, or Container Service instances for parallel access.

6. Click OK.

7.3.6 Mount NAS instances

After you create a NAS instance and add a mount point to it, you can mount the NAS instance to a compute node such as an ECS node.

Prerequisites

The following conditions determine whether an ECS instance can access a NAS instance through a mount point:

- **If the network type of the mount point is VPC, you can mount the NAS instance only to the ECS instances that are in the same VPC as the mount point. In addition, the VPC IP address of each of the ECS instances must match the authorized IP address of a rule in the permission group bound to the mount point.**
- **If the network type of the mount point is classic network, you can mount the NAS instance only to ECS instances under the same account as the mount point. In addition, ensure that the authorized IP address of a rule in the permission group bound to the mount point matches the intranet IP address of the ECS instance.**

Before you can use NFS to mount a NAS instance, ensure that `nfs-utils` or `nfs-common` has been installed. If not, run the following command to install the software package:

- **CentOS:** `sudo yum install nfs-utils`
- **Ubuntu or Debian:** `sudo apt-get install nfs-common`

Context

NAS supports the NFSv3 and NFSv4 protocols. You can choose a protocol version for mounting a NAS instance based on your scenario.

Use NFSv4.0 to mount a NAS instance

Format

```
sudo mount -t nfs -o vers=4.0 <domain name of the mount point>:<NAS
 instance directory> <target local mounting directory>
```

Parameter description

- **Domain name of the mount point: It is automatically generated when you create a NAS instance and a mount point.**
- **NAS instance directory: It is a directory of the NAS instance, which may be the root directory "/" or any subdirectory.**

- **Target local mounting directory: It is a directory on the local server, to which the NAS instance is to be mounted.**

Examples

- **Run the following command to mount the root directory of the NAS instance:**

```
mount -t nfs -o vers=4.0 014544bbf6-wdt41.regionid.nas.example.com
:/ /local/mntdir
```

- **Run the following command to mount the subdirectory named sub1 of the NAS instance:**

```
mount -t nfs -o vers=4.0 014544bbf6-wdt41.regionid.nas.example.com:/
sub1 /local/mntdir
```

Use NFSv3 to mount a NAS instance

Format

```
sudo mount -t nfs -o vers=3,nolock,proto=tcp <domain name of the mount
 point>:<NAS instance directory> <target local mounting directory>
```

Examples

- **Run the following command to mount the root directory of the NAS instance:**

```
mount -t nfs -o vers=3,nolock,proto=tcp 014544bbf6-wdt41.regionid.
nas.example.com:/ /local/mntdir
```

- **Run the following command to mount the subdirectory named sub1 of the NAS instance:**

```
mount -t nfs -o vers=3,nolock,proto=tcp 014544bbf6-wdt41.regionid.
nas.example.com:/sub1 /local/mntdir
```

View the mount point information

After the directories are mounted, run the following command to check the mounted NAS instance:

```
mount -l
```

Run the following command to check the current capacity of the mounted NAS instance:

```
df -h
```

7.4 NAS instance

7.4.1 View the NAS instance details

You can view details of an existing NAS instance in the NAS console, including system details and mount points.

Prerequisites

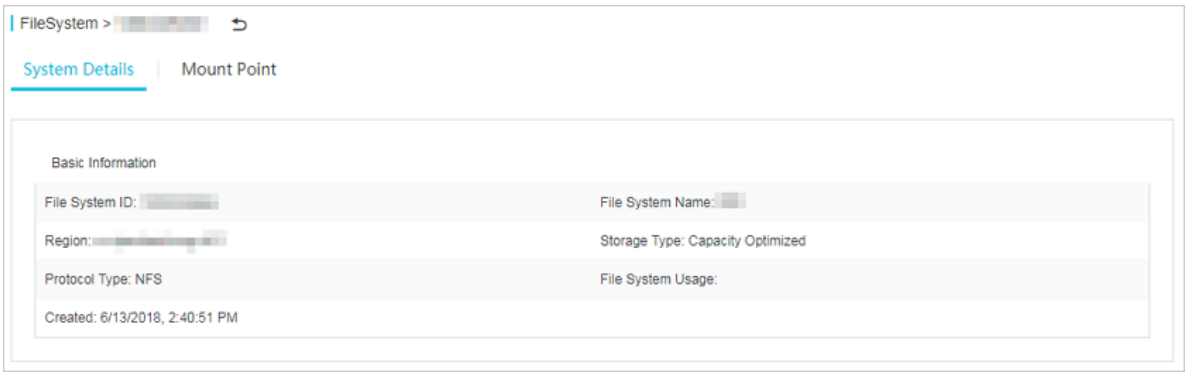
Before you can view NAS instance details, you must complete the procedure in *Quick start*, or ensure that at least one NAS instance has been created.

Procedure

1. *Log on to the NAS console*.
2. **On the File Storage NAS page, click the ID of the NAS instance that you want to view or click**  **> Details in the Actions column corresponding to the NAS**

instance ID. The NAS instance details page is displayed, as shown in *Figure 7-1: NAS instance details*.

Figure 7-1: NAS instance details



The NAS instance details page has two tabs:

- **The System Details tab displays the basic information about the NAS instance, including the NAS instance ID, region, and storage capacity.**
- **The Mount Point tab lists the mount points of the NAS instance. You can manage the mount points on this tab.**

7.4.2 Delete NAS instances

You can delete a NAS instance in the NAS console.

Prerequisites

Before you can delete a NAS instance, you must complete the procedure in *Quick start*, **or ensure that at least one NAS instance has been created and the NAS instance does not contain mount points.**

Procedure

1. *Log on to the NAS console*.
2. **On the File Storage NAS page, click**  **> Delete in the Actions column corresponding to the NAS instance that you want to delete.**
3. **In the Delete File System Instance message that appears, click OK.**

7.5 Mount point

7.5.1 View the mount point list

You can view a list of existing mount points in the NAS console.

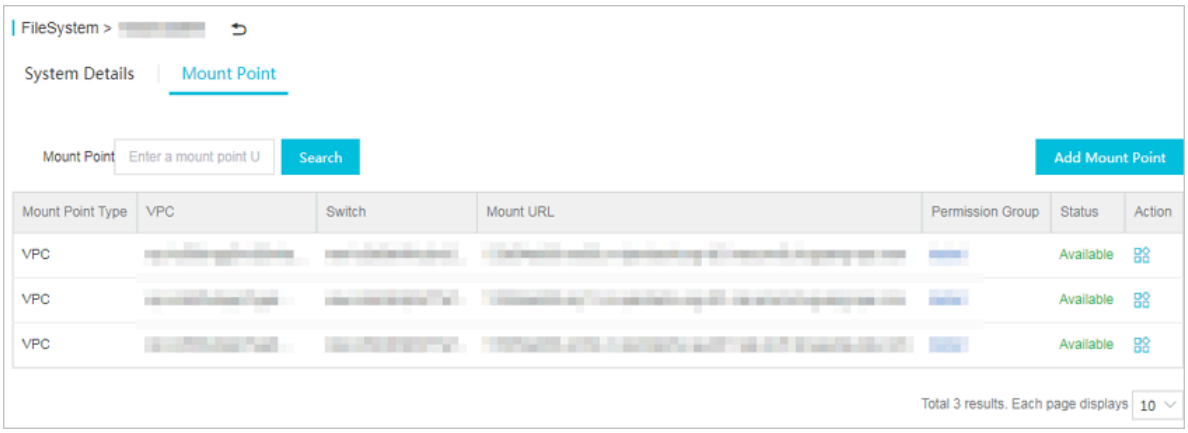
Prerequisites

Before you can view the mount point list, you must complete the procedure in *Quick start*, or ensure that at least one NAS instance and one mount point have been created.

Procedure

1. ***Log on to the NAS console*.**
2. **On the File Storage NAS page, click the ID of the NAS instance where the mount point that you want to view is located.**
3. **On the instance details page that appears, click the Mount Point tab. On the Mount Point tab that appears, you can view a list of all mount points in the NAS instance, as shown in *Figure 7-2: Mount point list*.**

Figure 7-2: Mount point list



7.5.2 Enable or disable mount points

You can enable or disable a mount point in the NAS console.

Prerequisites

Before you can enable or disable a mount point, you must complete the procedure in *Quick start*, or ensure that at least one NAS instance and one mount point have been created.

Procedure

1. *Log on to the NAS console*.
2. On the File Storage NAS page, click the ID of the NAS instance where the mount point that you want to view is located. On the instance details page that appears, click the Mount Point tab.
3. In the mount point list, you can perform the following operations:
   - Click  > Disable in the Actions column corresponding to the mount point that you want to disable. In the message that appears, click OK to disable access to the mount point from clients.
   - Click  > Enable in the Actions column corresponding to the mount point that you want to enable. In the message that appears, click OK to enable access to the mount point from clients.

7.5.3 Delete mount points

You can delete a mount point in the NAS console.

Prerequisites

Before you can delete a mount point, you must complete the procedure in *Quick start*, or ensure that at least one NAS instance and one mount point have been created.

Procedure

1. *Log on to the NAS console*.
2. On the File Storage NAS page, click the ID of the NAS instance where the mount point that you want to view is located. On the instance details page that appears, click the Mount Point tab.
3. In the mount point list, click  > Delete in the Actions column corresponding to the mount point that you want to delete.
4. In the Delete Mount Point message that appears, click OK.

> **Note:**
> After a mount point is deleted, it cannot be restored. Use caution when you delete a mount point.

7.5.4 Modify the permission group of a mount point

You must bind a permission group to each mount point. You can change the permission group that is bound to a mount point in the NAS console.

Prerequisites

Before you can change the permission group that is bound to a mount point, you must complete the procedure in *Quick start*, or ensure that at least one NAS instance and one mount point have been created in the region, and the mount point has been bound with a permission group.

Context

You must bind a permission group to each mount point. You can configure a source IP address list for the permission group to restrict access from ECS instances to the mount point. You can change the permission group that is bound to a mount point as required.

Procedure

1. ***Log on to the NAS console*.**
2. **On the File Storage NAS page, click the ID of the NAS instance where the mount point that you want to change is located. On the instance details page that appears, click the Mount Point tab.**
3. **In the mount point list, click  > Modify Permission Group in the Actions column of the mount point of which the permission group is to be modified.**
4. **In the Modify Mount Point Permission Group dialog box that appears, set Change to and click OK.**

> **Note:**
> **The modification may require up to 1 minute to take effect.**

7.6 Permission group

7.6.1 View the permission group list

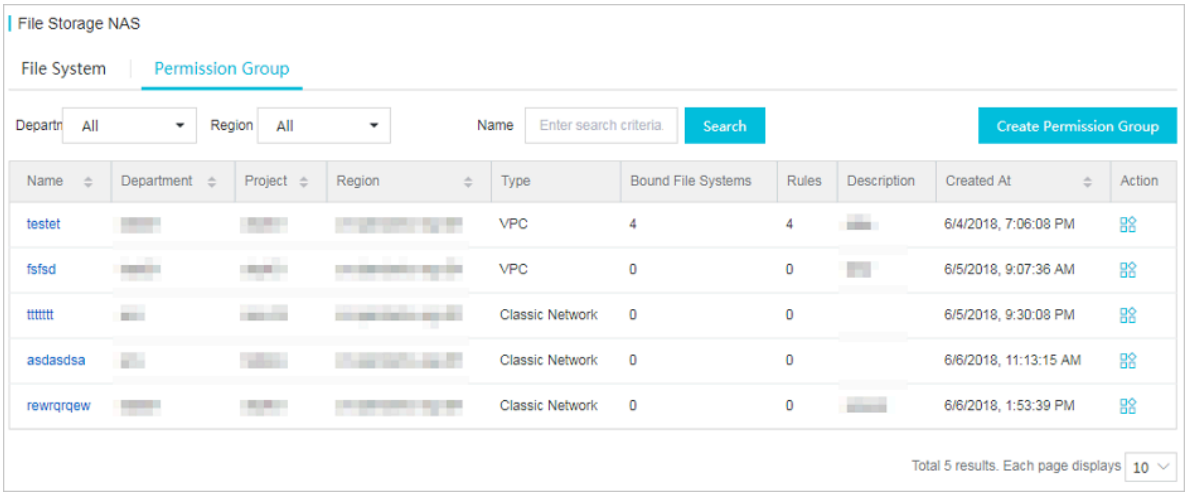
You can view a list of existing permission groups in the NAS console.

Prerequisites

Before you can view the permission group list, you must complete the procedure in *Quick start*, or ensure that at least one NAS instance and one permission group have been created.

Procedure

1. ***Log on to the NAS console*.**
2. **On the File Storage NAS page, click the Permission Group tab. On the Permission Group tab that appears, you can view a list of permission groups in the current region.**



7.6.2 Delete permission groups

You can delete a permission group in the NAS console.

Prerequisites

Before you can delete a permission group, you must complete the procedure in *Quick start*, or ensure that at least one NAS instance and one permission group have been created.

Procedure

1. ***Log on to the NAS console*.**
2. **On the File Storage NAS page, click the Permission Group tab.**
3. **In the permission group list, click > Delete in the Actions column corresponding to the permission group that you want to delete.**

4. In the Delete Permission Group message that appears, click OK.

> **Note:**
> Permission groups that are in use cannot be deleted. To delete a permission group in use, you must first disable it.

7.6.3 Manage permission group rules

You can manage the rules of a permission group in the NAS console, including modifying and deleting rules.

Prerequisites

Before you can manage the rules of a permission group, you must complete the procedure in *Quick start*, or ensure that at least one NAS instance and one permission group have been created, and the permission group has at least one rule.

Procedure

1. *Log on to the NAS console*.
2. On the File Storage NAS page, click the Permission Group tab.
3. On the Permission Group tab that appears, click the name of a permission group to go to the Rules List page.
4. On this page, you can modify or delete the rules of the permission group.
   - To modify a rule, click  > Change in the Actions column corresponding to the rule. In the Modify Rule dialog box that appears, modify Authorization Address, Read/Write Permissions, User Permission, or Priority. Click OK.
   - To delete a rule, click  > Delete in the Actions column corresponding to the rule. In the Delete Rule message that appears, click OK.

7.7 Migrate data

7.7.1 Migration tool for Windows

You can run the Network Attached Storage (NAS) migration tool for Windows after you download and extract it. With the migration tool, you can migrate data from on-premises disks or Object Storage Service (OSS) to a NAS file system.

Context

The migration tool provides the following features:

- **Supports data sources, such as on-premises disks, OSS, third-party cloud data stores, and HTTP repositories.**
- **Supports synchronizing stored data. You can specify a point in time and only migrate data with the last modification time later than the point in time.**
- **Supports synchronizing incremental data.**
- **Supports resumable upload and download.**
- **Supports the upload and download of data in parallel.**

Prerequisites

You must run the migration tool on an ECS instance on which the target NAS file system is applicable to mount. For more information about whether a NAS file system can be mounted on an ECS instance and how to mount a NAS file system, see *Mount a file system***.**

Supported operating systems

- **Windows Server 2008 Standard Service Pack 2 (SP2) 32-bit**
- **Windows Server 2008 R2 Datacenter 64-bit**
- **Windows Server 2012 R2 Datacenter 64-bit**
- **Windows Server 2016 Datacenter 64-bit**

Install and configure the migration tool

1. **Download the** *nasimport toolkit***.**
2. **Create a working directory on your local server for synchronization. For example , C:\NasImport. Then, extract the nasimport toolkit to the directory.**

3. **Edit the `config/sys.properties` configuration file in this directory.**

   We recommend that you use the default value of each configuration. You can change the value of a parameter based on your requirement. For more information see *List of parameters*.

   Table 7-4: Field description

Field	Description
workingDir	The working directory to which the nasimport toolkit is decompressed.
slaveTaskThreadNum	The number of working threads that run synchronization simultaneously.
slaveMaxThroughput(KB/s)	The upper limit of migration traffic.
slaveAbortWhenUncatchedException	Whether to skip an unknown error or abort . The process skips an unknown error by default.
dispatcherThreadNum	The number of parallel threads in a dispatching task. Keep the default value.

Start the nasimport service

Nasimport commands

- **Submit a job:** `nasimport -c config/sys.properties submit <your-job-configuration>`
- **Cancel a job:** `nasimport -c config/sys.properties clean <job-name>`
- **View a job:** `nasimport -c config/sys.properties stat detail`
- **Retry a job:** `nasimport -c config/sys.properties retry <job-name>`
- **Start the nasimport service:** `nasimport -c config/sys.properties start`

1. **Start the nasimport service.**

   Open the command prompt and switch to the working directory. Run the following command in the command prompt.

   ```
   nasimport -c config/sys.properties start
   ```

   Figure 7-3: Start the nasimport service

   Note:

   - **Ensure that the nasimport service is running. You can also set the nasimport service as a Windows background service.**
   - **When starting the nasimport service, run the following command to redirect the startup log of the service to a file named nasimport.log for later use.**

     ```
     nasimport -c config\sys.properties start > nasimport.log
         2>&1
     ```

2. **Define a job.**

   The config\local_job.cfg file is a template for defining jobs. You can use the template to define your jobs.

   Table 7-5: Field description

Field	Description
jobName	The name that uniquely identifies the job. You can submit multiple jobs with different names.

Field	Description
jobType	The job type. Values: import and audit. Import synchronizes data while audit checks the source and target data for consistency.
isIncremental=false	Whether to enable the automatic incremental mode. If this field is set to true, incremental data is scanned at the interval specified by incrementalModeInterval (in seconds) and synchronized to the NAS instance.
incrementalModeInterval=86400	The synchronization interval in the incremental mode. Unit: second.
importSince	The start time. Incremental data that is generated on and after this time point is synchronized to the NAS instance. This parameter is in the UNIX timestamp format. Unit: second. Default value: 0.
srcType	The synchronization source type. You can synchronize local files, files stored in OSS, or files stored in third-party cloud storage.
srcAccessKey	The AccessKey ID of the data source. Specify this field if you have set srcType to OSS or to a third-party cloud storage.
srcSecretKey	The AccessKey Secret of the data source. Specify this field if you have set srcType to OSS or to a third-party cloud storage.
srcDomain	The endpoint of the data source. **Note:** If the data source of a migration job is OSS, set srcDomain to the intranet domain name with "internal." With this setting, you can save the cost on downloading data from OSS and enjoy a faster migration service. You only pay for accessing OSS. You can obtain the intranet domain name of OSS in the OSS console. If your NAS instance is in a VPC and the data source is OSS, set srcDomain to the VPC domain name provided by OSS.

Field	Description
srcBucket	The source bucket name.
srcPrefix	The source prefix. Default value: null. If you have set srcType to local, enter the local directory to be synchronized. Note that the directory must be a full path ended with a forward slash (/). If you have set srcType to OSS or to a third-party cloud storage, enter the prefix of the object to be synchronized. To synchronize all files, set the prefix to null.
destType	The synchronization target type. Default value: NAS.
destMountDir	The local directory to which the NAS instance is mounted.
destMountTarget	The domain name of the NAS instance mount point.
destNeedMount=true	Whether nasimport performs automatic mounting. Default value: true. You can set this field to false and manually change the NAS instance mount point to the destMountDir directory.
destPrefix	The prefix of the synchronization target file. Default value: null.
taskObjectCountLimit	The maximum number of files that are processed by each task. This field affects the maximum number of parallel threads. It is usually set to the total number of files divided by the number of download threads that you have set. If you do not know the total number of files, you can keep the default value.
taskObjectSizeLimit	The maximum volume of the data downloaded by each task. Unit: byte.
scanThreadCount	The number of threads that scan files in parallel. This field affects file scan efficiency.

Field	Description
maxMultiThreadScanDepth	The maximum allowable depth of the directory in parallel scan. You can keep the default value.

Note:

- If you have configured the automatic incremental mode, the job runs periodically and permanently to scan the latest data.
- If you have set srcType to a third-party cloud storage, the List operation on files cannot implement checkpoints due to the API limits of third-party cloud storage. Killing the process before the List operation is complete may cause all the files to be relisted.

3. Submit a job.

   For example, you can migrate data stored in the *C:\Program Files\Internet Explorer* directory to NAS.

   a. Create a job. Make a copy of the *config\local_job.cfg* file and move the copy to the working directory. Then, modify the following fields of the copy.

srcType	local
srcPrefix	C:\\Program Files\\Internet Explorer
destMountDir	h:
destNeedMount	true
destMountTarget	xxxx-yyy.cn-beijing.nas.aliyuncs.com

Note:
The destMountDir field indicates the destination mount drive. The drive letter for the destination mount drive must be unique. The destMountTarget field indicates a NAS mount point.

   b. Submit a job. Open the command prompt and switch to the working directory. Then, run the `nasimport -c config\sys.properties submit local_job.cfg` command in the command prompt.

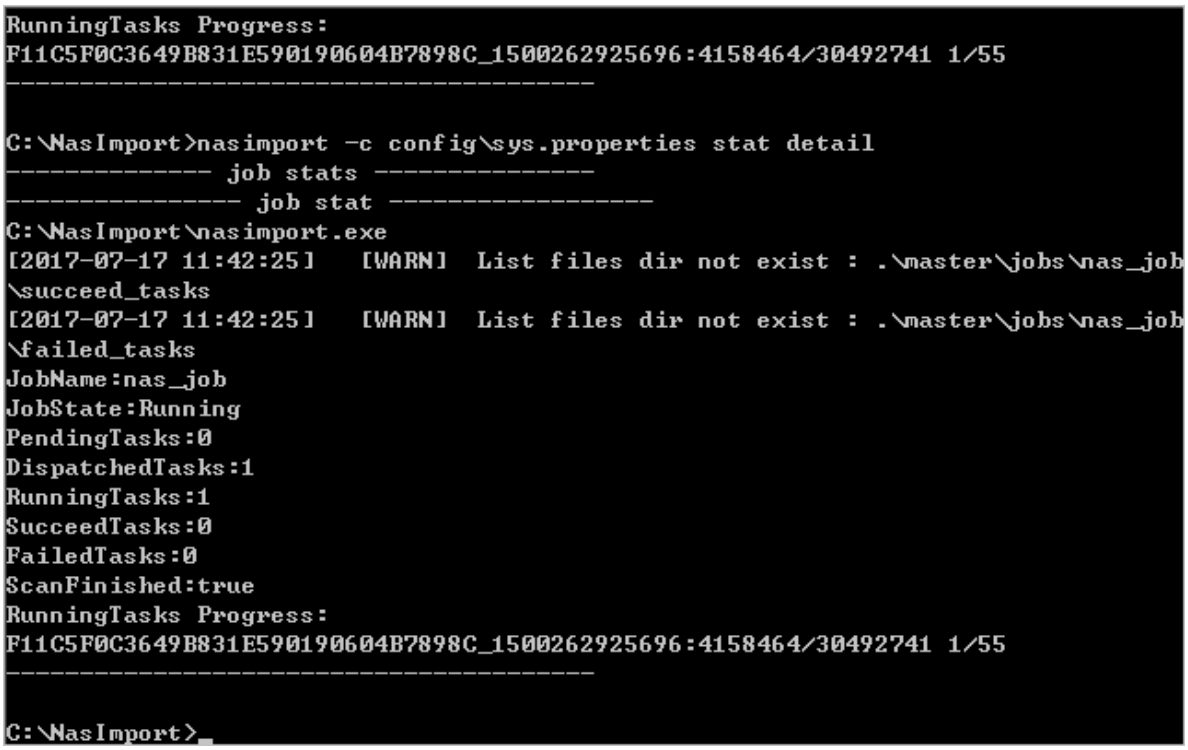
Note:

- If the name of the new job to be submitted is the same as that of a running job, you fail to submit the new job.

- You can stop the nasimport process to pause a synchronization job. When you need to synchronize data, you can restart the nasimport process to resume the synchronization job from the last break-point.

4. View job status. Run the following command.

```
nasimport -c config\sys.properties stat detail
```



The following figure shows the progress of the running job and the progress of each task. In this preceding figure, `4158464/30492741` indicates the size of uploaded data, which is 4,158,464 bytes, and the total size of data to be uploaded, which is 30,492,741 bytes. `1/1` indicates the total number of files, which is one, and the number of uploaded files, which is also one, respectively.

The migration tool separates a job you submit into multiple tasks and then runs these tasks. After all tasks are complete, the job is complete. After the job is complete, the `Succeed` or `Failed` status is displayed in the JobState field, which

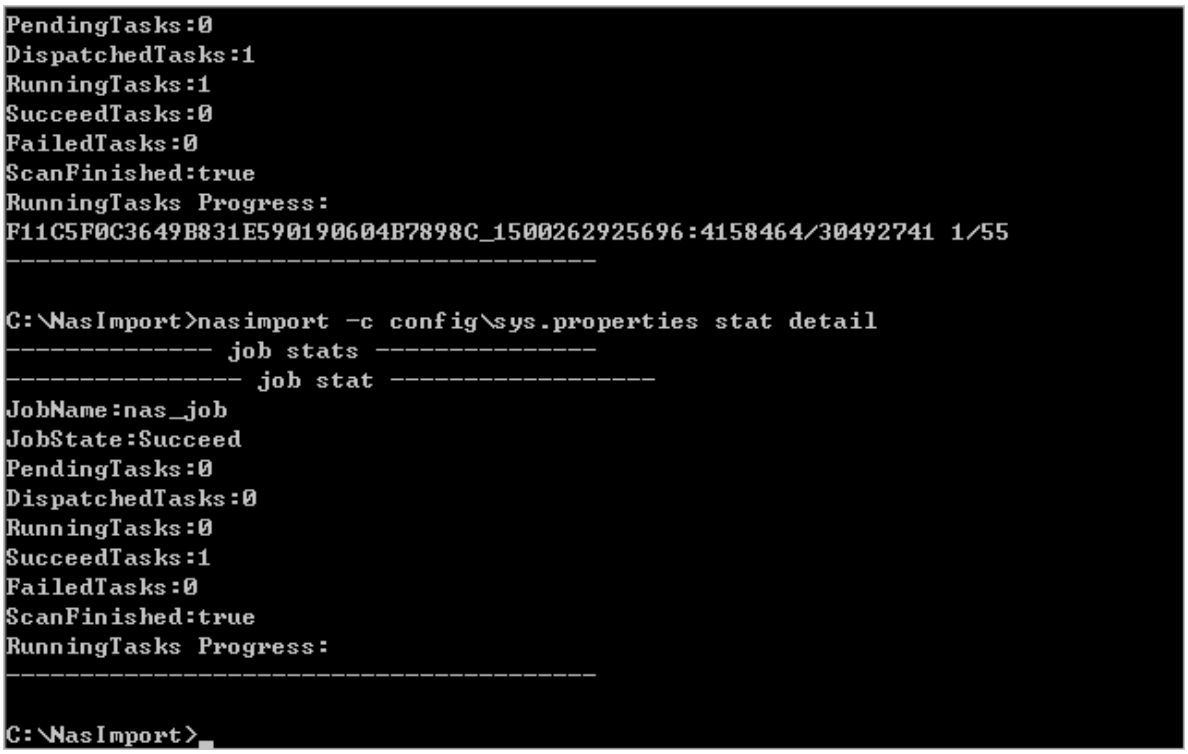
indicates whether the job is successful or failed. If a job fails, you can view the cause of the failure for each task in the following file.

```
master/jobs/$jobName/failed_tasks/*/audit.log
```

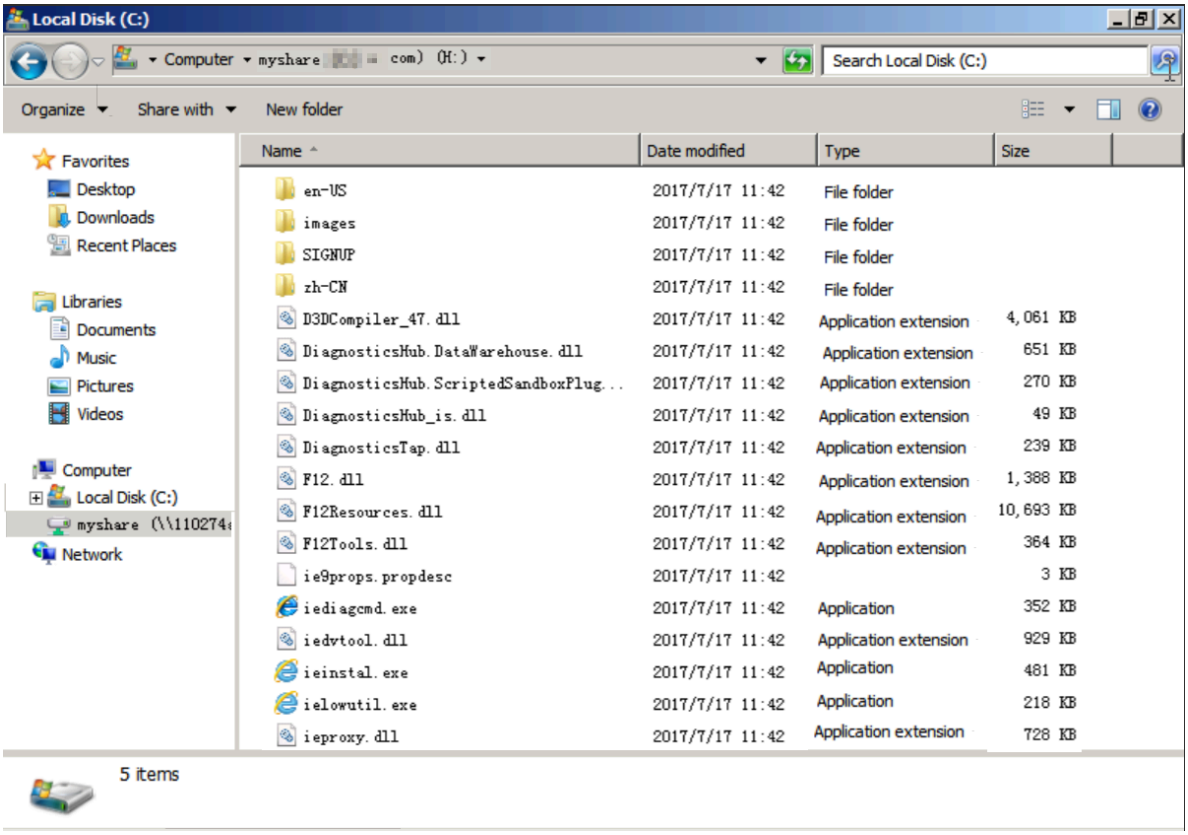
For failed tasks, a retry scheme is designed and built into the migration tool. For task failures due to unavailable data sources or target sources, you can run the following command to retry each failed task.

```
nasimport  -c config/sys.properties retry <job-name>
```

We recommend that you run the start detail command after a few minutes.



The value of the `SucceededTasks` changes to `1`, which indicates that a task is complete. Open a file manager, and you can find that files to be migrated already exist on the `H:` drive.



Common causes of failures

- **An error occurs during job configuration, such as an invalid AccessKey pair or AccessKey ID and insufficient permissions. This type of errors causes the failure of each task. You can open the nasimport.log file in the working directory to view the details of each failure. You must specify this file as the repository of logs for**

the migration tool when you start the nasimport service. You can also run the nasimport start command and view the results in the command prompt.

```
C:\NasImport\nasimport.exe
[2017-07-17 12:22:40]   [INFO]  JobDispatcher:Init
[2017-07-17 12:22:40]   [INFO]  job controller daemon start, working dir:.\
[2017-07-17 12:22:40]   [INFO]  watching job queue:.\master\jobqueue\
[2017-07-17 12:22:40]   [INFO]  JobDispatcher:Run
[2017-07-17 12:22:40]   [INFO]  try lock .\master\jobs\nas_job\.lock succeed
[2017-07-17 12:22:40]   [INFO]  start job:nas_job
[2017-07-17 12:22:40]   [INFO]  list checkpoint: .\master\jobs\nas_job\checkpoints\0, cpt
[2017-07-17 12:22:40]   [INFO]  scan task load checkpoint: [totalSize=0, totalCount=0, pre
[2017-07-17 12:22:40]   [INFO]  single thread scan start: nas_job
com.aliyun.oss.OSSException: The OSS Access Key Id you provided does not exist in our reco
        at com.aliyun.oss.common.utils.ExceptionFactory.createOSSException(ExceptionFactor
        at com.aliyun.oss.internal.OSSErrorResponseHandler.handle(OSSErrorResponseHandler.
        at com.aliyun.oss.common.comm.ServiceClient.handleResponse(ServiceClient.java:248)
        at com.aliyun.oss.common.comm.ServiceClient.sendRequestImpl(ServiceClient.java:130
        at com.aliyun.oss.common.comm.ServiceClient.sendRequest(ServiceClient.java:68)
        at com.aliyun.oss.internal.OSSOperation.send(OSSOperation.java:94)
        at com.aliyun.oss.internal.OSSOperation.doOperation(OSSOperation.java:149)
        at com.aliyun.oss.internal.OSSOperation.doOperation(OSSOperation.java:113)
        at com.aliyun.oss.internal.OSSBucketOperation.listObjects(OSSBucketOperation.java:
        at com.aliyun.oss.OSSClient.listObjects(OSSClient.java:526)
        at com.aliyun.ossimport2.master.scanner.OssLister.list(OssScanner.java:65)
        at com.aliyun.ossimport2.master.scanner.SingleThreadTask.run(SingleThreadTask.java
```

- **The encoding standard for the source data address is inconsistent with that for the destination data address. For example, GBK is implemented in Windows, while UTF-8 is implemented in Linux. For NAS data sources, this type of errors is prone to occur.**
- **A `SIZE_NOT_MATCH` error occurs in the audit.log file due to the change of a source file during migration. In this case, the original file is uploaded but the update of the change is not synchronized to NAS.**
- **An error occurs when you download a file because the file is deleted during upload.**
- **An error occurs when you download files from a data source because the data source is unavailable.**
- **An error may occur when you run the migration tool because you cancel a job before terminating the nasimport process.**
- **The migration tool does not stop running as expected. The status of the job is Abort. We recommend that you contact Alibaba Cloud Technical Support.**

7.7.2 Migrate local files or files stored in OSS to NAS instances

The nasimport tool helps you synchronize files and data from your local data center, OSS, or third-party cloud storage to NAS instances.

Context

Functions of nasimport:

- **Synchronizes local files, files stored in OSS, files in third-party cloud storage, or HTTP-linked files to NAS instances.**
- **Mounts NAS instances automatically.**
- **Synchronizes stored data (files modified after a specified time point).**
- **Synchronizes incremental data automatically.**
- **Supports resumable data transfer.**
- **Lists, uploads, and downloads data in parallel.**

To migrate a large volume of data (exceeding 2 TB) to a NAS instance in a short period of time, you can contact Alibaba Cloud technical support for a parallel synchronization on multiple machines in addition to using nasimport.

Runtime environment

Run the nasimport tool on an ECS virtual machine (VM) where you can mount the desired NAS instance. To determine whether the NAS instance can be mounted to an ECS VM and how to mount the NAS instance, see *Mount a NAS instance*.

You must run nasimport in Java JDK 1.7 or later. We recommend the *Oracle version JDK* **.**

> **Note:**
> **Before you start nasimport, run the `ulimit -n` command to view the number of files that the process allows you to open. If the number is smaller than 10,240, you must modify the configuration first.**

Deployment and configuration

1. **Create a working directory for synchronization on your local server, and download the nasimport toolkit to this directory.**

   Example: Run the following commands to create `/root/ms` as the working directory and download the toolkit to this directory:

   ```
   export work_dir=/root/ms
   wget http://docs-aliyun.cn-hangzhou.oss.aliyun-inc.com/assets/attach
   /45306/cn_zh/1479113980204/nasimport_linux.tgz
   tar zxvf ./nasimport_linux.tgz -C "$work_dir"
   ```

2. **Run the following commands to edit the configuration file named config/sys.properties in the working directory named $work_dir:**

   ```
   vim $work_dir/config/sys.properties
   workingDir=/root/ms
   ```

```
slaveUserName=
slavePassword=
privateKeyFile=
slaveTaskThreadNum=60
slaveMaxThroughput(KB/s)=100000000
slaveAbortWhenUncatchedException=false
dispatcherThreadNum=5
```

We recommend that you use the default configurations. When necessary, you can edit the configuration fields. For more information, see *Table 7-6: Field description*.

Table 7-6: Field description

Field	Description
workingDir	The working directory to which the nasimport toolkit is decompressed.
slaveTaskThreadNum	The number of working threads that run synchronization simultaneously.
slaveMaxThroughput(KB/s)	The upper limit of migration traffic.
slaveAbortWhenUncatchedException	Whether to skip an unknown error or abort. The process skips an unknown error by default.
dispatcherThreadNum	The number of parallel threads in a dispatching task. Keep the default value.

Running

The nasimport tool supports the following commands:

- **Submit a job:**

```
java -jar $work_dir/nasimport.jar -c $work_dir/config/sys.properties
 submit $jobConfigPath
```

- **Cancel a job:**

```
java -jar $work_dir/nasimport.jar -c $work_dir/config/sys.properties
 clean $jobName
```

- **View job status:**

```
```

```
java -jar $work_dir/nasimport.jar -c $work_dir/config/sys.properties
 stat detail
```

- **Retry the job:**

```
java -jar $work_dir/nasimport.jar -c $work_dir/config/sys.properties
 retry $jobName
```

Perform the following procedure to run a migration job:

1. **Run the following commands to start nasimport:**

```
cd $work_dir
nohup java -Dskip_exist_file=false -jar $work_dir/nasimport.jar -c $
work_dir/config/sys.properties start > $work_dir/nasimport.log 2>&1
 &
```

Note:

The related log file is automatically generated in the directory where nasimport is started. We recommend that you start nasimport in the working directory named $work_dir. If the value of `skip_exist_file` is true when you start nasimport, nasimport skips the files that already exist in the NAS instance with the length the same as the source.

2. **Edit the sample job description file named *`nas_job.cfg`*.**

Table 7-7: Field description

Field	Description
jobName	The name that uniquely identifies the job. You can submit multiple jobs with different names.
jobType	The job type. Values: import and audit. Import synchronizes data while audit checks the source and target data for consistency.
isIncremental=false	Whether to enable the automatic incremental mode. If this field is set to true, incremental data is scanned at the interval specified by incrementa lModeInterval (in seconds) and synchronized to the NAS instance.
incrementalModeInterval= 86400	The synchronization interval in the incremental mode. Unit: second.

Field	Description
importSince	The start time. Incremental data that is generated on and after this time point is synchronized to the NAS instance. This parameter is in the UNIX timestamp format. Unit: second. Default value: 0.
srcType	The synchronization source type. You can synchronize local files, files stored in OSS, or files stored in third-party cloud storage.
srcAccessKey	The AccessKey ID of the data source. Specify this field if you have set srcType to OSS or to a third-party cloud storage.
srcSecretKey	The AccessKey Secret of the data source. Specify this field if you have set srcType to OSS or to a third-party cloud storage.
srcDomain	The endpoint of the data source. **Note:** If the data source of a migration job is OSS, set srcDomain to the intranet domain name with "internal." With this setting, you can save the cost on downloading data from OSS and enjoy a faster migration service. You only pay for accessing OSS. You can obtain the intranet domain name of OSS in the OSS console. If your NAS instance is in a VPC and the data source is OSS, set srcDomain to the VPC domain name provided by OSS.
srcBucket	The source bucket name.

Field	Description
srcPrefix	The source prefix. Default value: null. If you have set srcType to local, enter the local directory to be synchronized. Note that the directory must be a full path ended with a forward slash (/). If you have set srcType to OSS or to a third-party cloud storage, enter the prefix of the object to be synchronized. To synchronize all files, set the prefix to null.
destType	The synchronization target type. Default value: NAS.
destMountDir	The local directory to which the NAS instance is mounted.
destMountTarget	The domain name of the NAS instance mount point.
destNeedMount=true	Whether nasimport performs automatic mounting. Default value: true. You can set this field to false and manually change the NAS instance mount point to the destMountDir directory.
destPrefix	The prefix of the synchronization target file. Default value: null.
taskObjectCountLimit	The maximum number of files that are processed by each task. This field affects the maximum number of parallel threads. It is usually set to the total number of files divided by the number of download threads that you have set. If you do not know the total number of files, you can keep the default value.
taskObjectSizeLimit	The maximum volume of the data downloaded by each task. Unit: byte.
scanThreadCount	The number of threads that scan files in parallel. This field affects file scan efficiency.

Field	Description
maxMultiThreadScanDepth	The maximum allowable depth of the directory in parallel scan. You can keep the default value.

Note:

- If you have configured the automatic incremental mode, the job runs periodically and permanently to scan the latest data.
- If you have set srcType to a third-party cloud storage, the List operation on files cannot implement checkpoints due to the API limits of third-party cloud storage. Killing the process before the List operation is complete may cause all the files to be relisted.

3. Submit the job.

```
java -jar $work_dir/nasimport.jar -c $work_dir/config/sys.properties
  submit $work_dir/nas_job.cfg
```

Note:

- If the job that you want to submit has the same name as a job in progress, you cannot submit the job.
- To pause a synchronization job, stop the nasimport process. You can restart the nasimport process to resume synchronization from where it was paused.
- To resynchronize all files, stop the nasimport process and run the following command to clear the current job. For example, the job name is nas_job (you can set the job name in the nas_job.cfg file).

```
ps axu | grep "nasimport.jar.* start" | grep -v grep | awk '{
print "kill -9 "$2}' | bash
java -jar $work_dir/nasimport.jar  -c $work_dir/conf/sys.
properties clean nas_job
```

4. Check the job status.

```
java -jar $work_dir/nasimport.jar -c $work_dir/config/sys.properties
  stat detail
--------------job stats begin---------------
----------------job stat begin------------------
JobName:nas_job
JobState:Running
PendingTasks:0
RunningTasks:1
SucceedTasks:0
FailedTasks:0
ScanFinished:true
RunningTasks Progress:
```

```
FD813E8B93F55E67A843DBCFA3FAF5B6_1449307162636:26378979/26378979 1/1
----------------job stat end-----------------
--------------job stats end---------------
```

The preceding command output displays the overall progress of the current job and the progress of the current task. For example, `26378979/26378979` indicates that the total volume of data to be uploaded is 26,378,979 bytes and the volume of data already uploaded is 26,378,979 bytes. `1/1` indicates that the total number of files to be uploaded is 1 and the number of files already uploaded is 1.

The migration tool splits each job that you submit into multiple tasks for parallel execution. After all the tasks are complete, the job is considered complete. After a job is complete, JobState displays `Succeed` or `Failed`, to indicate whether the job is successful or not. If a job fails, run the following command to check the failure cause of each task.

In the following command, replace $jobName with the actual job name (you can set jobName in the local_job.cfg file).

```
cat $work_dir/master/jobs/$jobName/failed_tasks/*/audit.log
```

We have already retried failed jobs in nasimport. If a failure is caused by the temporary unavailability of the source or target data, run the following command to retry the job:

```
java -jar $work_dir/nasimport.jar  -c $work_dir/config/sys.
properties retry $jobNam
```

Common causes of job failures

- **The job configurations are incorrect, for example, the AccessKey ID is incorrect or permissions are insufficient. In this case, all tasks fail. To identify this cause, check the *`$work_dir/nasimport.log`* file.**
- **The encoding method of source file names is inconsistent from the default file name encoding method of the system (GBK for Windows and UTF-8 for Linux). This is the typical cause of failure for NFS data sources.**
- **A file in the source directory is modified during the upload process. This cause is indicated by a `SIZE_NOT_MATCH` error in audit.log. In this case, the old file is uploaded, but the changes are not synchronized to the NAS instance.**
- **The source file is deleted during the upload process, causing file download to fail .**

- An error occurs in the data source, causing source data download to fail.
- The Clean operation is performed before the nasimport process is killed, which may cause a program execution error.
- The nasimport tool aborts and the job status is Abort. If this failure occurs, contact Alibaba Cloud technical support.

7.8 Directory-level ACL

NAS allows you to configure an Access Control List (ACL) for a directory to control access to the directory and files in it.

Prerequisites

- You must use the NFSv4 protocol to mount a NAS instance on a client.
- You must use the alinas-acl tool to configure an ACL. To ensure correct permission settings, do not change the mode or run the chmod command to modify the file permissions.

Procedure

1. Run the `sudo mount -t nfs -o vers=4.0 <mount point domain name>:<NAS directory> <target directory on the current server>` command, such as `mount -t nfs -o vers=4.0 014544bbf6-wdt41.cn-hangzhou.nas.aliyuncs.com:/ /mnt`, to ensure that a NAS instance has been mounted by using the NFSv4 protocol.

   Note:

   - The value of the vers parameter varies with the client version. If an error occurs when you set vers to 4.0, set vers to 4 instead.
   - If a NAS instance with the ACL feature disabled has been mounted before, we recommend that you mount the instance again to ensure that the ACL feature takes effect.

2. Run the following command to install `nfs4-acl-tools` in CentOS:

   ```
   sudo yum install nfs4-acl-tools -y
   ```

3. Run the following command to ensure that Python 2.7.5 has been installed:

```
python --version Python 2.7.5
```

4. **Use alinas-acl to make ACL settings.**

```
./alinas_acl set ./foo --add --user Alice --rule r #Grant the read
permission on the foo file to user Alice.
./alinas_acl set ./foo -a -u Alice -r r #Abbreviated format of the
previous command
./alinas_acl set ./dir --add --group Staff --rule rwx #Grant the
read, write, and execute permissions on the dir directory to group
Staff.
./alinas_acl set ./foo --add --user EVERYONE@ --rule none #Grant no
permissions to user EVERYONE@.
./alinas_acl set ./foo --add --user 1001 --rule none #Grant no
permissions to the user whose uid is 1001.
./alinas_acl set ./dir -d -u Bob #Revoke the permissions of user Bob
 on the dir directory.
```

Note:

- **To avoid performance deterioration, we recommend that you configure an ACL for a directory, instead of for files in the directory.**
- **The number of Access Control Entries (ACEs) for a single file must not exceed 10.**

5. **View the ACL.**

```
./alinas_acl get ./foo #View the permissions on the foo file. # file
: foo/ # owner:
                root # group: root OWNER@::rw- GROUP@::r-- EVERYONE
@:::--- Alice::r-- Staff:g:rwx
                1001::---
```

Note:

OWNER@, GROUP@, and EVERYONE@ are three special usernames that are automatically generated when you configure the ACL. They correspond to the user, group, and others classes in the mode operand. If there are conflicts between the permissions specified in the ACL and the mode operand, the actual permissions may vary based on the client version.

8 Apsara File Storage for HDFS

8.1 What is Apsara File Storage for HDFS?

Apsara File Storage for HDFS is a file storage service for computing resources such as Alibaba Cloud ECS instances and Container Service.

It supports standard HDFS access protocols. You can use Apsara File Storage for HDFS without modifying the existing big data analytics applications. Apsara File Storage for HDFS offers various features such as unlimited capacity, performanc e expansion, single namespace, multi-party sharing, high reliability, and high availability. Compared with user-created HDFS storage, Apsara File Storage for HDFS greatly reduces maintenance costs and data security risks.

You can perform the following operations on Apsara File Storage for HDFS:

- **Create a file system instance and a mount point.**
- **Create permission groups for Apsara File Storage for HDFS instances and create rules for permission groups to allow access or grant different levels of access permissions to IP addresses or IP address segments.**
- **Access file system instances through standard HDFS protocol interfaces within computing resources such as ECS and Container Service.**
- **Perform basic and advanced operations on file systems, mount points, and permission groups in the Apsara File Storage for HDFS console.**
- **Perform basic and advanced operations on Apsara File Storage for HDFS through SDKs or APIs.**

8.2 Limits

Before using Apsara File Storage for HDFS, you need to understand the following service limits.

Hadoop FileSystem or AbstractFileSystem

- **Does not support the settings of directory modification time (mtime) and access time (atime), or settings of file mtime and atime attributes through setTimes.**
- **Does not support symbolic links.**
- **Does not support file truncation (truncate).**

- **Does not support file concatenation (concat).**
- **Does not support extended attributes (XAttrs).**
- **Does not support snapshot operations.**
- **Does not support delegation token operations.**
- **Does not support checksum operations such as setWriteChecksum and setVerifyChecksum.**
- **Does not support ACL operations.**
- **Does not support file block locations.**

Hadoop fs command line tool

- **Does not support snapshot commands (createSnapshot, deleteSnapshot, and renameSnapshot).**
- **Does not support ACL commands (setfacl and getfacl).**
- **Does not support XAttr commands (setfattr and getfattr).**
- **Does not support file truncation commands (truncate).**

8.3 Quick start

8.3.1 Log on to the Apsara File Storage for HDFS console

This topic describes how to log on to the Apsara File Storage for HDFS console.

Prerequisites

- **Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://***IP address or domain name of the Apsara Stack console***/manage.**
- **We recommend that you use the Chrome browser.**

Procedure

1. **Open your browser.**
2. **In the address bar, enter the access address of the Apsara Stack console in the format of http://***IP address or domain name of the Apsara Stack console***/manage, and then press Enter.**

3. Enter the correct username and password.
   - The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click LOGIN to go to the Dashboard page.
5. In the top navigation bar, choose  > Compute, Storage & Networking > Apsara File Storage for HDFS.

8.3.2 Create a file system

Apsara File Storage for HDFS runs as file system instances. Before using Apsara File Storage for HDFS, you must create a file system instance.

Procedure

1. *Log on to the Apsara File Storage for HDFS console*.
2. On the File Systems tab of the Apsara File Storage for HDFS page, click Create File System.

   Note:

   - You can create up to 1,000 file systems.
   - The upper limit of the file system capacity is 10 PB.

   To raise the limit, contact the administrator.
3. On the Create HDFS File System page, configure the required parameters.

Category	Parameter	Description
Region	Region	Select a region.
	Zone	Select a zone from the drop-down list.

Category	Parameter	Description
Basic Settings	Department	Select a department from the drop-down list.
	Project	Select a project from the drop-down list.
	File System Name	Specify a name for the file system. The file system name cannot be empty. The name can be a maximum of 100 bytes in length and must be globally unique.
	Description	Enter the description of the file system.
Storage Configuration	Protocol Type	Select HDFS.
	Storage Type	Select STANDARD.
	File System Capacity (GB)	Enter the capacity of the file system. Note: The capacity ranges from 1 GB to 10 TB.

4. Click OK.

   If the newly created file system does not appear on the File Systems tab, refresh the page.

8.3.3 Create a permission group

Apsara File Storage for HDFS manages permissions on file systems through permission groups. Before using the file system, you need to create a permission group and configure relevant parameters.

Context

In Apsara File Storage for HDFS, the permission group acts as a whitelist. You can create rules for the permission group to allow specified IP addresses or IP address segments to access the file system, and assign different levels of access permissions to different IP addresses or IP address segments.

Procedure

1. *Log on to the Apsara File Storage for HDFS console*.
2. On the Permission Groups tab of the Apsara File Storage for HDFS page, click Create Permission Group.

   Note:
   You can create up to 100 permission groups. To raise the limit, contact the administrator.
3. In the Create Permission Group dialog box that appears, set the required parameters.

Parameter	Description
Region	Select a region.
Department	Select a department from the drop-down list.
Project	Select a project from the drop-down list.
Name	Specify the name of the permission group. The permission group name cannot be empty. It can be 6 to 100 characters in length. It can contain letters, digits, and hyphens (-). The name can be a maximum of 100 bytes in length and must be globally unique.
Network Type	Select Classic Network or VPC.
Description	Enter the description of the permission group.

4. Click OK.

8.3.4 Create a permission group rule

A permission group in Apsara File Storage for HDFS is a group of rules, which enable you to manage permissions of file system instances. Before using Apsara

File Storage for HDFS, you need to create rules and set parameters for the created permission group.

Context

In Apsara File Storage for HDFS, the permission group acts as a whitelist. You can create rules for the permission group to allow specified IP addresses or IP address segments to access the file system, and assign different levels of access permissions to different IP addresses or IP address segments.

 **Warning:**
To ensure the security of your data, we recommend that you exercise caution when creating permission group rules and granting permissions to IP addresses.

Procedure

1. *Log on to the Apsara File Storage for HDFS console*.
2. On the Permission Groups tab of the Apsara File Storage for HDFS page, click a Permission Group ID from the list.

    **Note:**
   You can create up to 1,000 permission group rules. To raise the limit, contact the administrator.
3. Click Create Rule in the upper-right corner.
4. In the Create Rule dialog box that appears, set the required parameters.

Parameter	Description
Access Type	The optional value is Readable and Writable, allowing authorized objects to read from and write to the file system.
Authorized IP Address	The address can be an IP address or IP address segment, which indicates the authorized object of this rule, such as 192.168.1.2 or 192.168.1.0/24.
Priority	The priority value ranges from 1 to 100. The value 1 indicates the highest priority. When an authorized object matches multiple rules, the rule with the highest priority takes effect.
Description	Enter the description of the permission group rule.

5. Click OK to create the permission group rule.

8.3.5 Create a mount points

After a file system and its permission group are created, you need to create mount points to mount the file system to compute nodes (ECS or Container Service instances).

Context

A mount point is the access address of a file system instance in a VPC or classic network. Each mount point corresponds to a domain name. Apsara File Storage for HDFS only supports VPC mount points.

Procedure

1. *Log on to the Apsara File Storage for HDFS console*.
2. On the File Systems tab of the Apsara File Storage for HDFS page, click a File System ID.
3. On the Mount Point tab, click Add Mount Point.

   Note:
   You can create up to 100 mount points. To raise the limit, contact the administrator.

4. In the Add Mount Point dialog box that appears, configure the required parameters.

Parameter	Description
Region	The region of the file system for which you want to create a mount point.
File System	The file system for which you want to create a mount point.
Network Type	Select Classic Network or VPC. If the network type is VPC, you also need to set VPC and VSwitch.
Permission Group	Select the permission group bound to the mount point .

Parameter	Description
Description	Add the description of the mount point.

 **Note:**

- **Ensure that a VPC and a VSwitch have been created.**
- **You can mount a mount point on multiple compute nodes (ECS or Container Service instances) for shared access.**
- **To create multiple mount points for one file system, ensure that the configuration of each mount point is unique. At least one item among the permission group, VPC, or VSwitch of a mount point must be different from the other mount points.**

5. **Click OK.**

8.3.6 Mount a file system

After creating a file system and a mount point, you can mount Apsara File Storage for HDFS instances through the mount point.

The file system of Apsara File Storage for HDFS supports Hadoop 2.7.x protocols.

Prerequisites

The following conditions determine whether an ECS instance can access a file system through a mount point:

- **If the network type of the mount point is VPC, you can mount the file system on only ECS instances in the same VPC. In addition, ensure that the authorized IP address of a rule in the permission group bound to the mount point matches the VPC IP address of the ECS instance.**
- **Apsara File Storage for HDFS provides a UserGroupService API based on Linux `/etc/passwd`. For more information, see *UserGroupService*.**

 **Note:**

- **If no changes have been made to the configurations, you must ensure that the content of the `/etc/passwd` file is consistent across all compute nodes to guarantee permission control over files and directories.**

- **You can achieve customized user or group information management through the UserGroupService interface and configure `alidfs.usergroupservice.impl` in *`core-site.xml`* to access SDKs of Apsara File Storage for HDFS.**

- **Before mounting a file system through HDFS, ensure that Java 1.8 is installed on the ECS instance.**

UserGroupService

In Hadoop, user and group information associated with files and directories are stored as strings. In Apsara File Storage for HDFS, user and group information associated with files and directories are stored as 32-bit integers. When you create a file or directory in the Apsara File Storage for HDFS SDK, the user information obtained through UserGroupInformation is converted into a UID, and the group information obtained through UserGroupService is converted into a GID. When you obtain file or directory information, the UID is converted into a user name, and the GID is converted into a group name.

Apsara File Storage for HDFS provides the UserGroupService interface, which enables you to:

- **Maintain mappings between users and groups.**
- **Maintain mappings between UserNames and UIDs.**
- **Maintain mappings between GroupNames and GIDs.**

URI format

The URI in Apsara File Storage for HDFS must be in the dfs://DfsInstanceID.RegionID.alidfs.aliyun.com:10290 format. Example: dfs://f-63a47d43wh98.cn-neimeng-env10-d01.alidfs.aliyun.com:10290.

Procedure

The following steps show you how to mount a file system in Apsara File Storage for HDFS.

1. **Configure core-site.xml.**

   Add the following content to the *`core-site.xml`* file on a node and synchronize the file content to all nodes dependent on hadoop-common:

```
<property>
     <name>fs.defaultFS</name>
     <value>dfs://DfsInstanceID.RegionID.alidfs.aliyun.com:10290</
value>
```

```
</property>
<property>
     <name>fs.dfs.impl</name>
     <value>com.alibaba.dfs.DistributedFileSystem</value>
</property>
<property>
     <name>fs.AbstractFileSystem.dfs.impl</name>
     <value>com.alibaba.dfs.DFS</value>
</property>
<property>
<name>alidfs.usergroupservice.impl</name>
<value>com.alibaba.dfs.security.LinuxUserGroupService.class</value>
</property>
```

Note:

- **Replace RegionID and DfsInstanceID with the actual region ID and Apsara File Storage for HDFS instance ID.**
- **You must synchronize the content in *core-site.xml* to all nodes dependent on hadoop-common.**

2. **Deploy the Apsara File Storage for HDFS SDK.**

   Download *alicloud.dfs-1.0.0.jar* and deploy it on the CLASSPATH of the Hadoop ecosystem component.

   Note:
   We recommend that you deploy it in the directory where hadoop-common-X.YZ.jar is located.

   The following figure shows the directory structure of Spark 2.3.0 after decompression:

```
[root@Hadoop3 spark-2.3.0-bin-hadoop2.7]# ls
bin   data       examples  kubernetes  licenses  metastore_db  python  README.md  sbin  yarn
conf  derby.log  jars      LICENSE     logs      NOTICE        R       RELEASE    work
```

   You need to copy *alicloud.dfs-1.0.0.jar* to the *jars* directory.

3. **Optimize the configuration.**

   The Apsara File Storage for HDFS SDK provides configuration items such as `io.file.buffer.size` and `dfs.connection.count`, to optimize application performance. After you configure the items in *core-site.xml*, you must synchronize the file content to all nodes dependent on hadoop-common.

```
<property>
     <name>io.file.buffer.size</name>
```

```
      <value>4194304</value>
      <description>To achieve high throughput, no less than 1MB, no
more than 8MB</description>
</property>
<property>
      <name>dfs.connection.count</name>
      <value>1</value>
      <description>If multi threads in the same process will read/
write to DFS, set to count of threads</description>
</property>
```

4. **Verify the installation.**

   After the deployment and configuration are complete, run the `hadoop fs-ls/` **command for verification.**

```
[hadoop@iZ5wf05xt7fvxpnkx15oy2Z ~/hadoop-2.7.2]$ bin/hadoop fs -ls /
Found 12 items
drw-r----T   - hadoop       hadoop         75498848 1970-01-01 08:00 /MR
drwxr----T   - alicloud-dfs alicloud-dfs   75498848 1970-01-01 08:00 /benchmarks
drw-r----T   - hadoop       hadoop         75498848 1970-01-01 08:00 /hadoop
drwxr----T   - hadoop       hadoop         75498848 1970-01-01 08:00 /tcpds
drw-r----T   - alicloud-dfs alicloud-dfs   75498848 1970-01-01 08:00 /tmp
```

   If no error is reported, the file system is successfully mounted.

8.4 File systems

8.4.1 View file system details

You can view the details of a file system in the Apsara File Storage for HDFS console.

Prerequisites

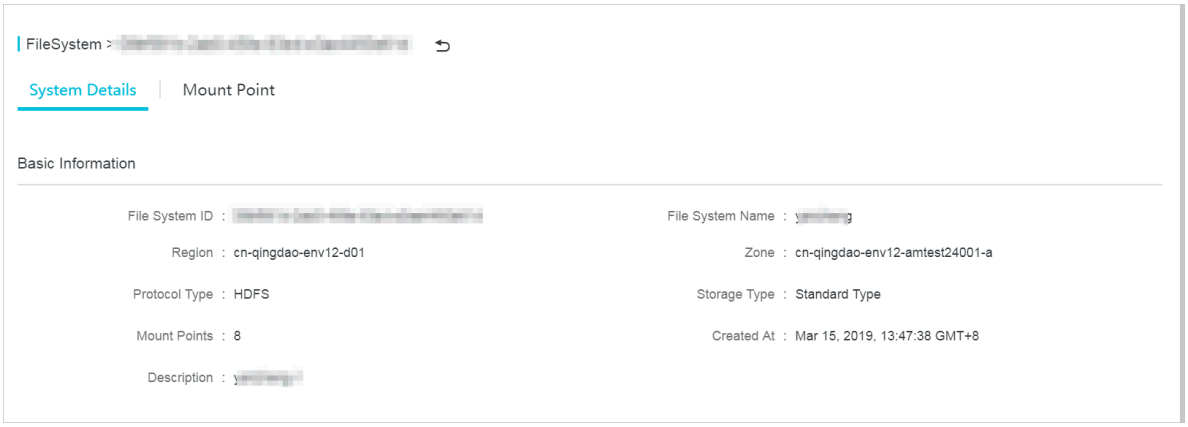
Before viewing file system details, you need to complete the steps in *Quick start*, **or ensure that at least one file system has been created in the region.**

Procedure

1. *Log on to the Apsara File Storage for HDFS console*.

2. On the Apsara File Storage for HDFS page, click a File System ID or click the icon in the Actions column corresponding to a file system and choose Details from the shortcut menu.

   The System Details page shows basic information of the file system, such as file system ID, region, and capacity information, as shown in the following figure.



8.4.2 Delete a file system

You can delete file systems in the Apsara File Storage for HDFS console.

Prerequisites

Before deleting a file system, you need to complete the steps in *Quick start*, or ensure that at least one file system has been created in the region.

Procedure

1. *Log on to the Apsara File Storage for HDFS console*.
2. Click the icon in the Actions column corresponding to a file system instance and choose Delete from the shortcut menu.
3. In the Delete File System Instance message that appears, click OK.

8.4.3 Modify file system information

You can modify file system information in the Apsara File Storage for HDFS console.

Prerequisites

Before modifying the information, you need to complete the steps in *Quick start*, or ensure that at least one file system and permission group have been created in the region.

Procedure

1. *Log on to the Apsara File Storage for HDFS console*.
2. Click the  icon in the Actions column corresponding to a file system instance and choose Change from the shortcut menu.
3. In the Change File System dialog box that appears, set Name, Description, and File System Capacity as needed.

   Note:
   The maximum capacity of a file system is 10 TB.
4. Click OK.

8.5 Mount points

8.5.1 View the list of mount points

You can view the list of mount points in the Apsara File Storage for HDFS console.

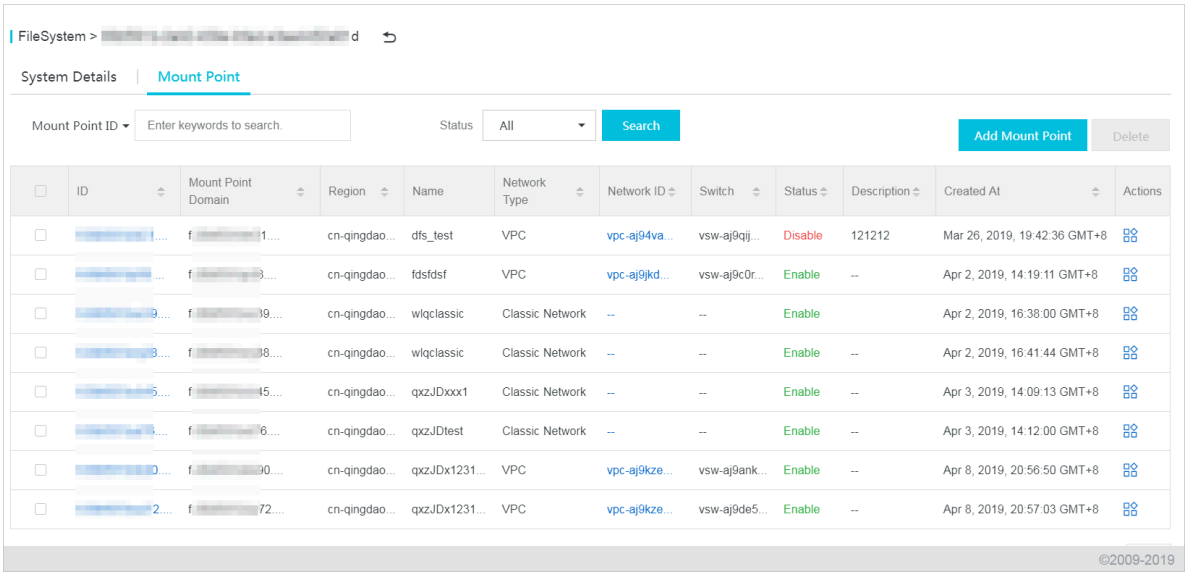
Prerequisites

Before viewing the list of mount points, you need to complete the steps in *Quick start*, or ensure that at least one file system and mount point have been created in the region.

Procedure

1. *Log on to the Apsara File Storage for HDFS console*.
2. On the File Systems tab of the Apsara File Storage for HDFS page, click a File System ID in the list.

3. Click the Mount Point tab to view the list of all mount points in the file system, as shown in the following figure.



8.5.2 Manage a mount point

You can modify, enable, or delete a mount point in the Apsara File Storage for HDFS console.

Prerequisites

Before modifying, enabling, or deleting a mount point, you need to complete the steps in *Quick start*, or ensure that at least one file system and one mount point have been created in the region.

Procedure

1. *Log on to the Apsara File Storage for HDFS console*.
2. On the Apsara File Storage for HDFS page, click a File System ID. On the file system details page that appears, click Mount Point.

3. You can modify, disable, and delete mount points.
   - Locate the mount point that you want to modify and perform the following operations:
     a. Click the  icon in the Actions column corresponding to the mount point and choose Change from the shortcut menu.
     b. In the Change Mount Point dialog box that appears, you can modify the permission group to which the mount point is bound, change the status to Enabled or Disable, or modify the description of the mount point.
     c. After you complete the modifications, click OK.
   - Locate the mount point that you want to disable and perform the following operations:
     a. Click the  icon in the Actions column corresponding to the mount point and choose Disable from the shortcut menu.
     b. In the message that appears, click OK.
   - Locate the mount point that you want to delete and perform the following operations:
     a. Click the  icon in the Actions column corresponding to the mount point and choose Delete from the shortcut menu.
     b. In the message that appears, click OK.

8.6 Permission groups

8.6.1 View the list of permission groups

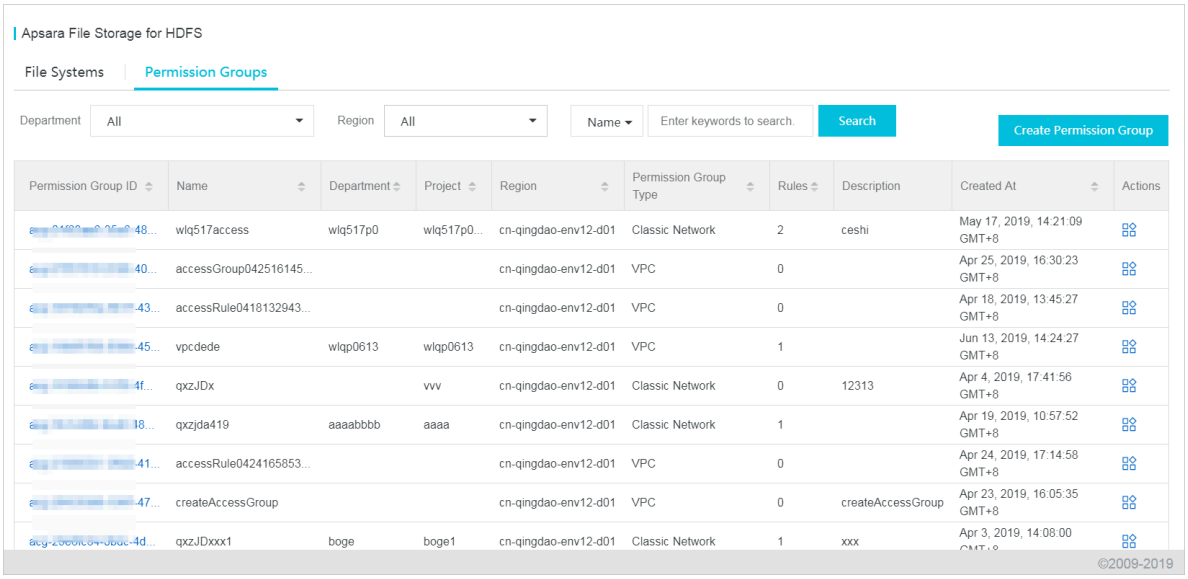
You can view the list of permission groups in the Apsara File Storage for HDFS console.

Prerequisites

Before viewing the list, you need to complete the steps in *Quick start*, or ensure that at least one file system and one permission group have been created in the region.

Procedure

1. *Log on to the Apsara File Storage for HDFS console*.
2. On the Apsara File Storage for HDFS page, click the Permission Groups tab to view the list of permission groups, as shown in the following figure.



8.6.2 Modify permission group information

You can modify the information about a permission group in the Apsara File Storage for HDFS console, including name and description.

Prerequisites

Before modifying the information, you need to complete the steps in *Quick start*, or ensure that at least one file system and permission group have been created in the region.

Context

You must bind a permission group to each of your mount points. Each permission group has a source IP address whitelist that is used to restrict access to the mount point from ECS instances. You can modify the permission group bound to a mount point as required.

Procedure

1. *Log on to the Apsara File Storage for HDFS console*.
2. On the Apsara File Storage for HDFS page, click the Permission Groups tab.
3. Click the icon in the Actions column corresponding to a permission group and choose Change from the shortcut menu.

4. In the Change Permission Group dialog box that appears, modify Name and Description as needed.
5. Click OK.

8.6.3 Delete a permission group

You can delete permission groups in the Apsara File Storage for HDFS console.

Prerequisites

Before deleting a permission group, you need to complete the steps in *Quick start*, or ensure that at least one file system and one permission group have been created in the region.

Procedure

1. *Log on to the Apsara File Storage for HDFS console*.
2. On the Apsara File Storage for HDFS page, click the Permission Groups tab.
3. Click the  icon in the Actions column corresponding to a permission group and choose Delete from the shortcut menu.
4. In the Delete Permission Group message that appears, click OK.

8.6.4 Manage a permission group rule

You can manage the rules of a permission group in the Apsara File Storage for HDFS console, including modifying and deleting rules.

Prerequisites

Before managing a permission group rule, you need to complete the steps in *Quick start* or ensure that at least one file system and permission group have been created in the region, and the permission group has at least one rule.

Context

In Apsara File Storage for HDFS, the permission group acts as a whitelist. You can create rules for the permission group to allow specified IP addresses or IP address segments to access the file system, and assign different levels of access permissions to different IP addresses or IP address segments.

 **Warning:**

To ensure the security of your data, we recommend that you exercise caution when creating permission group rules and granting permissions to IP addresses.

Procedure

1. *Log on to the Apsara File Storage for HDFS console*.
2. On the Apsara File Storage for HDFS page, click the Permission Groups tab.
3. Click the  icon in the Actions column corresponding to a permission group and choose Manage Rules from the shortcut menu.
4. On the rule list page, you can modify or delete a rule.

   Perform the following procedure to modify a rule of the permission group:

   a. Click the  icon in the Actions column corresponding to the rule and choose Change from the shortcut menu.
   b. In the Change Rule dialog box that appears, you can modify Access Type and Priority.
   c. Click OK.

   Perform the following procedure to delete a rule of the permission group:

   a. Click the  icon in the Actions column corresponding to the rule and choose Delete from the shortcut menu.
   b. In the Delete Rule dialog box that appears, click OK.

9 ApsaraDB for RDS

9.1 What is ApsaraDB for RDS?

ApsaraDB for RDS is a stable, reliable, and automatically scaling online database service.

Based on the distributed file system and high-performance storage, ApsaraDB for RDS allows you to easily perform database operations and maintenance with its complete set of solutions for disaster recovery, backup, restoration, monitoring, and migration.

ApsaraDB for RDS supports three storage engines: MySQL, PostgreSQL, and PPAS. These storage engines can help you create database instances suitable to your business needs.

ApsaraDB RDS for MySQL

Originally based on a branch of MySQL, ApsaraDB RDS for MySQL has proven its performance and throughput during the high-volume concurrent traffic of Double 11. ApsaraDB RDS for MySQL provides whitelist configuration, backup and restoration, transparent data encryption, data migration, and management for instances, accounts, and databases. It also provides the following advanced features:

- **Read-only instance: In scenarios where RDS has a small number of write requests but a large number of read requests, you can enable read/write splitting to distribute read requests away from the primary instance. Read-only instances allow ApsaraDB RDS for MySQL 5.6 to automatically scale the reading capability and increase the application throughput when a large amount of data is being read.**
- **Read/write splitting: The read/write splitting feature provides an extra read/write splitting endpoint. This endpoint enables an automatic link for the primary instance and all its read-only instances. An application can use this method to read and write data by connecting to the read/write splitting endpoint. Write requests are automatically distributed to the primary instance while read requests are distributed to read-only instances based on their**

weights. To scale up the reading capacity of the system, you can add more read-only instances.

- **Data compression: ApsaraDB RDS for MySQL 5.6 allows you to compress data by using the TokuDB storage engine. Data transferred from the InnoDB storage engine to the TokuDB storage engine can be reduced by 80% to 90% in volume. 2 TB of data in InnoDB can be compressed to 400 GB or less in TokuDB. In addition to data compression, TokuDB supports transaction and online DDL operations. TokuDB is compatible with MyISAM and InnoDB applications.**

ApsaraDB RDS for PostgreSQL

PostgreSQL is the most advanced open source database that is fully compatible with SQL and supports a diverse range of data formats such as JSON, IP, and geometric data. In addition to support for features such as transactions, subqueries, multi-version concurrency control (MVCC), and data integrity check, ApsaraDB RDS for PostgreSQL integrates a series of features including high availability, backup, and restoration to ease operations and maintenance loads.

ApsaraDB RDS for PostgreSQL provides basic features such as whitelist configurat ion, backup and restoration, data migration, and management for instances, accounts, and databases.

ApsaraDB RDS for PPAS

Postgres Plus Advanced Server (PPAS) is a stable, secure, and scalable enterprise -class relational database. Based on PostgreSQL, PPAS features enhanced performance, application solutions, and compatibility. It is able to directly run Oracle applications. You can run enterprise-class applications on PPAS in a stable and cost-effective manner.

ApsaraDB RDS for PPAS provides basic features such as whitelist configuration, backup and restoration, data migration, and management for instances, accounts, and databases.

9.2 Instructions

9.2.1 Limits on ApsaraDB RDS for MySQL

Before you use ApsaraDB RDS for MySQL, you must understand its limits and take precautions.

To guarantee instance stability and security, ApsaraDB RDS for MySQL has some service limits, as listed in *Table 9-1: Limits on ApsaraDB RDS for MySQL*.

Table 9-1: Limits on ApsaraDB RDS for MySQL

Operation	Description
Database parameter modification	Database parameters can only be modified through the RDS console or API operations. Due to security and stability considerations, only specific parameters can be modified.
Root permissions of databases	The root and SA permissions are not provided.
Database backup	· Logical backup can be performed through the command line interface (CLI) or graphical user interface (GUI). · Physical backup can only be performed through the RDS console or API operations.
Database restoration	· Logical restoration can be performed through the CLI or GUI. · Physical restoration can only be performed through the RDS console or API operations.
Data import	· Logical import can be performed through the CLI or GUI. · Data can only be imported through the MySQL CLI.
ApsaraDB RDS for MySQL storage engine	· Only InnoDB and TokuDB are supported. Due to the inherent shortcomings of the MyISAM engine, some data may be lost. Only some stock instances use the MylSAM engine. MyISAM engine tables in newly created instances will be automatically converted to InnoDB engine tables. · For safety performance and security considerations, we recommend that you use the InnoDB storage engine. · The Memory engine is not supported. Newly created Memory tables will be automatically converted into InnoDB tables.

Operation	Description
Database replication	ApsaraDB RDS for MySQL provides dual-node clusters based on a primary/secondary replication architecture. The secondary instances in this replication architecture are hidden and cannot be accessed directly.
RDS instance restart	Instances must be restarted through the RDS console or API operations.
Account and database management	ApsaraDB RDS for MySQL uses the RDS console to manage accounts and databases. ApsaraDB RDS for MySQL also allows you to create a privileged account to manage users, passwords, and databases.
Standard account	· Custom authorization is not supported. · The account management and database management interfaces are provided in the RDS console. · Instances that support standard accounts also support privileged accounts.
Privileged account	· Custom authorization is supported. · The RDS console does not provide interfaces to manage accounts or databases. Relevant operations can only be performed through code or DMS. · The privileged account cannot be reverted back to a standard account.

9.2.2 Usage limits of ApsaraDB RDS for PostgreSQL

Before you use ApsaraDB RDS for PostgreSQL, you need to understand its limits and take precautions against them.

To guarantee instance stability and security, ApsaraDB RDS for PostgreSQL has some service limits, as listed in *Table 9-2: Limits on ApsaraDB RDS for PostgreSQL*.

Table 9-2: Limits on ApsaraDB RDS for PostgreSQL

Operation	Description
Database parameter modification	Not supported.
Root permission of databases	Superuser permissions are not provided.

Operation	Description
Database backup	Data can only be backed up by using pg_dump.
Data migration	Only PostgreSQL can be used to restore data that was backed up by using pg_dump.
Database replication	· The system automatically builds HA databases based on PostgreSQL streaming replication without user input. · PostgreSQL standby nodes are hidden and cannot be accessed directly.
RDS instance restart	RDS instances must be restarted from the RDS console or through APIs.
Network settings	For instances that are operating in safe mode, net.ipv4.tcp_timestamps cannot be enabled in SNAT mode.

9.2.3 Usage limits of ApsaraDB RDS for PPAS

Before you use ApsaraDB RDS for PPAS, you must understand its limits and take precautions against them.

To guarantee instance stability and security, ApsaraDB RDS for PPAS has some service limits, as listed in *Table 9-3: Limits on ApsaraDB RDS for PPAS*.

Table 9-3: Limits on ApsaraDB RDS for PPAS

Operation	Description
Database parameter modification	Not supported.
Root permission of databases	Superuser permissions are not provided.
Database backup	Data can only be backed up by using pg_dump.
Data migration	Only PostgreSQL can be used to restore data that was backed up by using pg_dump.
Database replication	· The system automatically builds HA databases based on PPAS streaming replication without user input. · PPAS standby nodes are hidden and cannot be accessed directly.

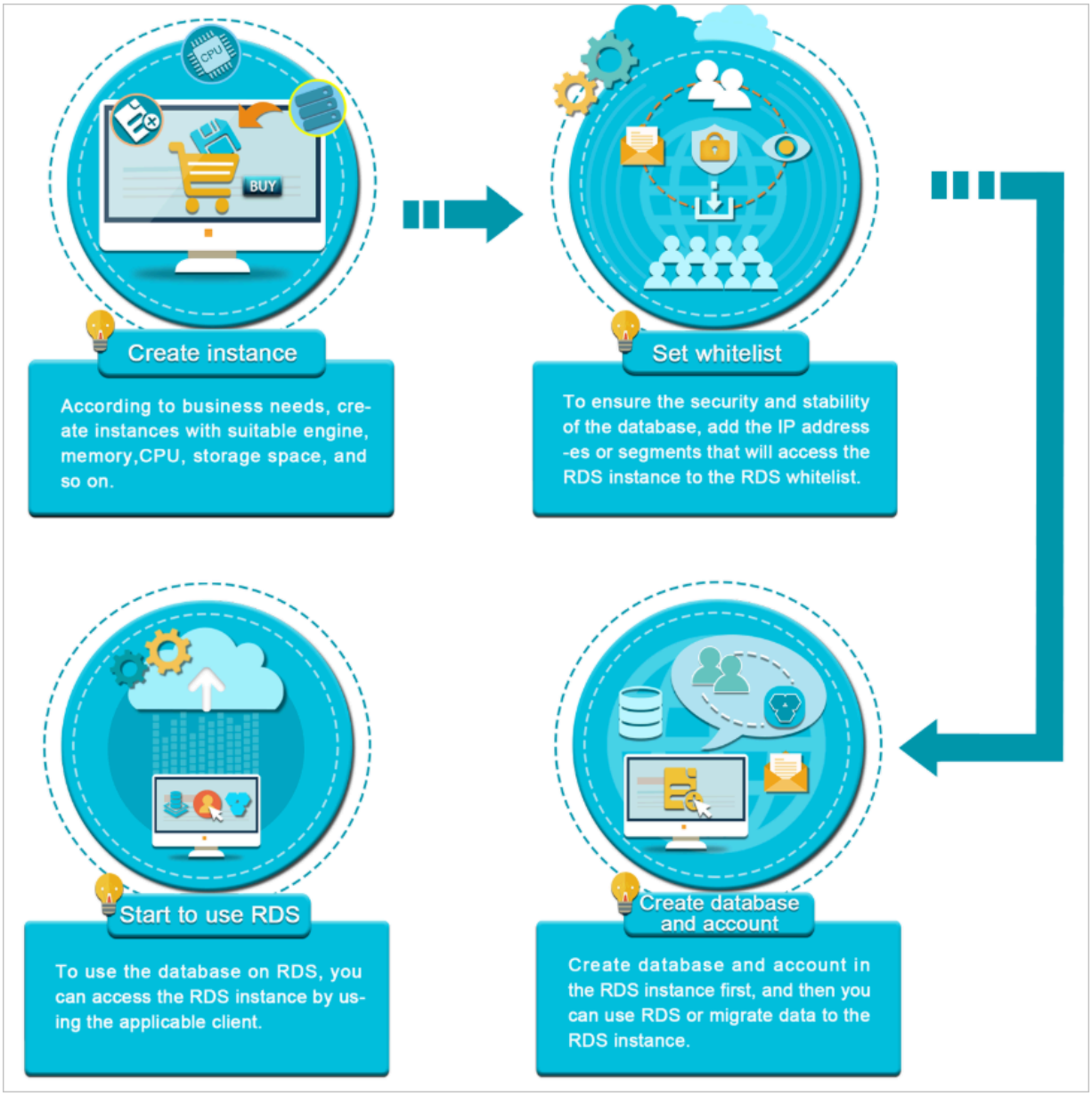
Operation	Description
RDS instance restart	RDS instances must be restarted from the RDS console or through APIs.
Network settings	For instances that are operating in safe mode, net.ipv4.tcp_timestamps cannot be enabled in SNAT mode.

9.3 Quick start

9.3.1 Quick start

ApsaraDB for RDS quick start covers the following topics: creating an RDS instance, configuring a whitelist, creating a database, creating an account, and connecting the instance. This topic uses ApsaraDB RDS for MySQL as an example to describe how to use RDS. It provides all the necessary information to build an RDS instance.

Typically, you must complete several operations after instance creation to make it ready for use, as shown in *Quick start flow*.



- *Create an instance*

  An instance is a virtualized database server on which you can create and manage multiple databases.

- *Configure a whitelist*

  After creating an RDS instance, you must configure its whitelist to allow access from external devices.

  The whitelist can improve the access security of your RDS instance. We recommend that you maintain the whitelist on a regular basis. The whitelist configuration process does not affect the normal operations of the RDS instance.

- *Create a database* and *account*

  Before you use a database, you must first create the database and an account in the RDS instance. Different engines support different account modes. For more information, see the console UI and documentation.

- *Log on to an instance through DMS* or connect to the instance from a client

  After creating an RDS instance, configuring a whitelist, and creating a database and an account, you can use Data Management Service (DMS) or a general database client to connect to the instance.

9.3.2 Log on to the RDS console

This topic describes how to log on to the RDS console.

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://*IP address or domain name of the Apsara Stack console*/manage.
- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).

4. Click LOGIN to go to the Dashboard page.
5. In the top navigation bar, choose > Database > Relational Database Service.

9.3.3 Create an instance

This topic describes how to create an instance in the RDS console.

Prerequisites

Before you create an RDS instance, you must apply for an Apsara Stack tenant account.

Procedure

1. *Log on to the RDS console*.
2. On the Relational Database Service (RDS) page, click Create Instance. On the Create Instance page that appears, configure the parameters as promoted.

   The parameters are described as follows.

   Table 9-4: Instance creation parameters

Category	Parameter	Description
Basic Settings	Department	The department to which the instance belongs.
	Project	The project to which the instance belongs.
	Region	The region where the instance is located. Services in different regions are not interconnected over the internal network. Once a region is selected, it cannot be changed.
	Zone	The zone of the instance. Common RDS instances use a hot standby architecture. Single zone indicates that the primary and secondary nodes are in the same zone.
Specifications	Instance Name	The name of the RDS instance. It must start with a letter . It can contain letters, digits, underscores (_), and hyphens (-). It must be 2 to 64 characters in length.
	Database Type	The database type varies with region. The available database types are displayed on the Create Instance page.
	Database Version	The version of the database.

Category	Parameter	Description
	CPU/ Memory	The specifications of the instance. Available specifications are as follows: · Dedicated: This type of specifications is followed by the "Dedicated" suffix. · Dedicated Hosts: This type of specifications is followed by the "Dedicated Hosts" suffix. Memory size determines the maximum number of connections and IOPS. The actual values are displayed on the console UI.
	Storage Capacity	The storage space of the instance, including the space for data, system files, binlog files, and transaction files.
Network Type	Instance Type	The type of the instance. Based on your business scenario, you can select from either internal instance or external instance. External instances can only support the classic network type.
	Network Type	RDS instances support the following network types: · `Classic Network`: Cloud services in a classic network are not isolated. Unauthorized access to a cloud service is only blocked by the security group or whitelist policy of the service. · `VPC`: A VPC helps you to build an isolated network environment in Alibaba Cloud. You can customize the routing table, IP address range, and gateway in a VPC. We recommend that you select VPC for enhanced security. You can create a VPC in advance, or change the network type to VPC after creating an instance.

Category	Parameter	Description
Connection Mode	Connection Mode	RDS instances support two connection modes: `Standard Mode` and `Safe Mode`. · Standard Mode: RDS uses SLB to eliminate the impact of database engine HA switching on the application layer. This mode shortens the response time, but slightly increases the probability of transient disconnections and disables SQL interception. · Safe Mode: This mode prevents 90% of transient disconnections and intercepts SQL injection attacks based on semantic analysis. However, it increases the response time by over 20%.
Quantity	Instances	The number of RDS instances that can be created simultaneously. Maximum value: 20.

3. After you set the parameters, click Create.

9.3.4 Configuration initialization

9.3.4.1 RDS for MySQL

9.3.4.1.1 Configure a whitelist

To guarantee database security and reliability, you must modify the whitelist of an RDS instance before you enable the instance. You must add the IP addresses or CIDR blocks that are used for database access to the whitelist of the RDS instance.

Context

The system creates a default whitelist group for each instance. This whitelist group can be modified or cleared, but cannot be deleted. For each newly created RDS instance, IP address 0.0.0.0/0 is added to the default whitelist group by default. 0.0.0.0/0 allows all IP addresses to access the instance, which greatly reduces database security. Delete 0.0.0.0/0 from the whitelist.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane of the Basic Information page, choose Security Control > Whitelist Settings.

4. You can use the following methods to add an IP address or CIDR block:

   - Add an IP address or CIDR block to the default whitelist group.

   a) Click the icon corresponding to the default whitelist group, and add an IP address or CIDR block.

   b) Click OK.

   - Create a whitelist group and add an IP address or CIDR block to the group.

   a) Click Create Whitelist Group. In the Create Whitelist Group dialog box that appears, set Group Name and IP Addresses.

   b) Click OK.

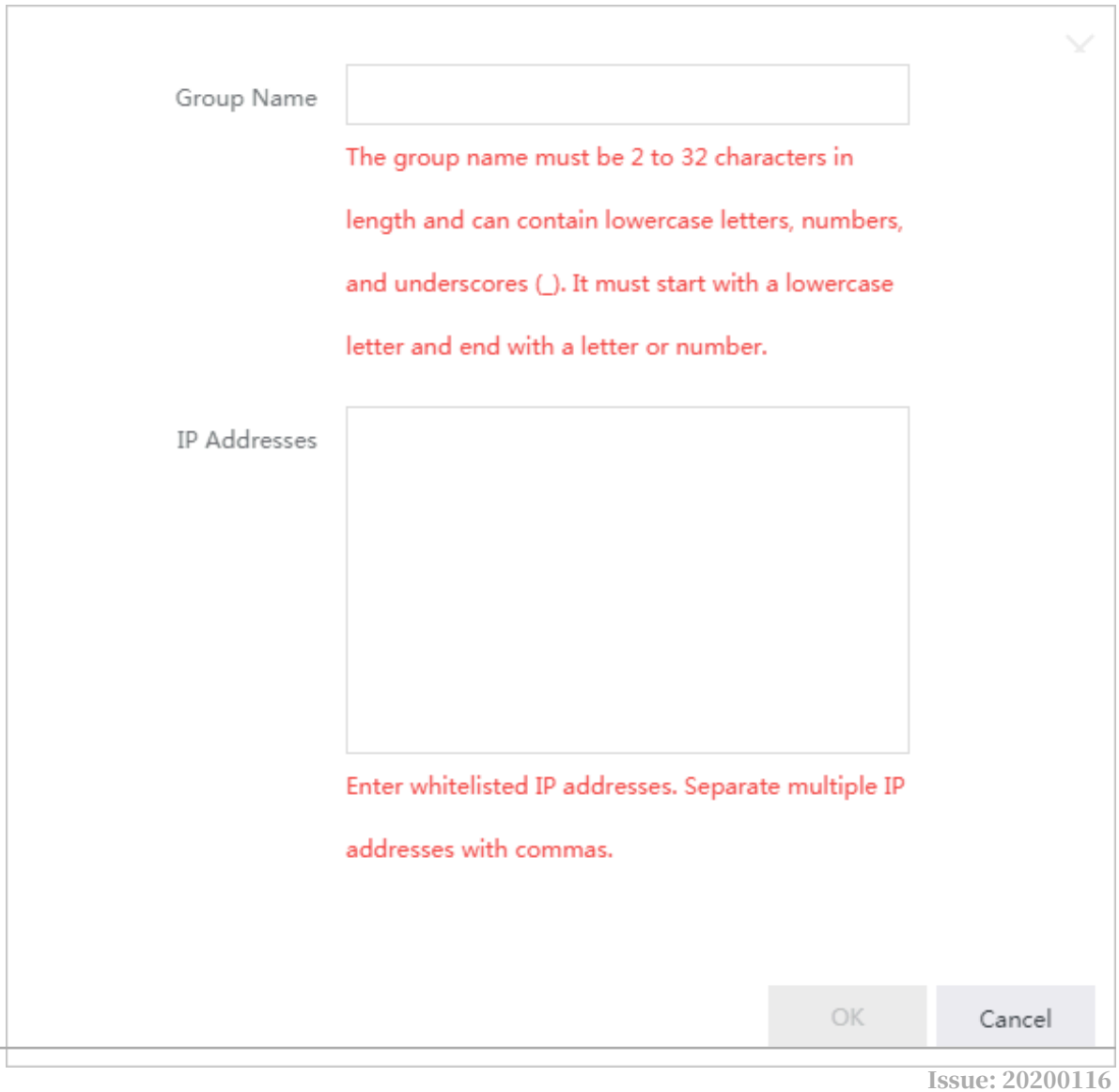
   Table 9-5: Whitelist configuration parameters describes the parameter configurations.

Table 9-5: Whitelist configuration parameters

Parameter	Description
Group Name	The name of the new whitelist group. The naming conventions are as follows: · It must start with a lowercase letter and end with a lowercase letter or digit. · It can contain lowercase letters, digits, and underscores (_). · It must be 2 to 32 characters in length. **Note:** You cannot change the name of an existing whitelist group.

Parameter	Description
IP Addresses	The IP addresses or CIDR blocks that are allowed to access the RDS instance. Note: · If you enter a CIDR block, such as 10.10.10.0/24, any IP addresses in the 10.10.10.X format can access the RDS instance. · If you enter multiple IP addresses, use commas (,) to separate them. Example: 192.168.0.1,172.16.213.9 · 127.0.0.1 indicates that no IP addresses are allowed to access the RDS instance. · 0.0.0.0/0 indicates that all IP addresses are allowed to access the RDS instance.

Figure 9-1: Create a whitelist group



Notice:

The proper use of the whitelist can improve the access security of your RDS instance. We recommend that you maintain the whitelist on a regular basis. After you configure the whitelist, you can perform the following operations:

- **Click the icon to modify the whitelist group.**
- **Click the icon to clear the default whitelist group or delete a custom whitelist group.**

9.3.4.1.2 Create a privileged account

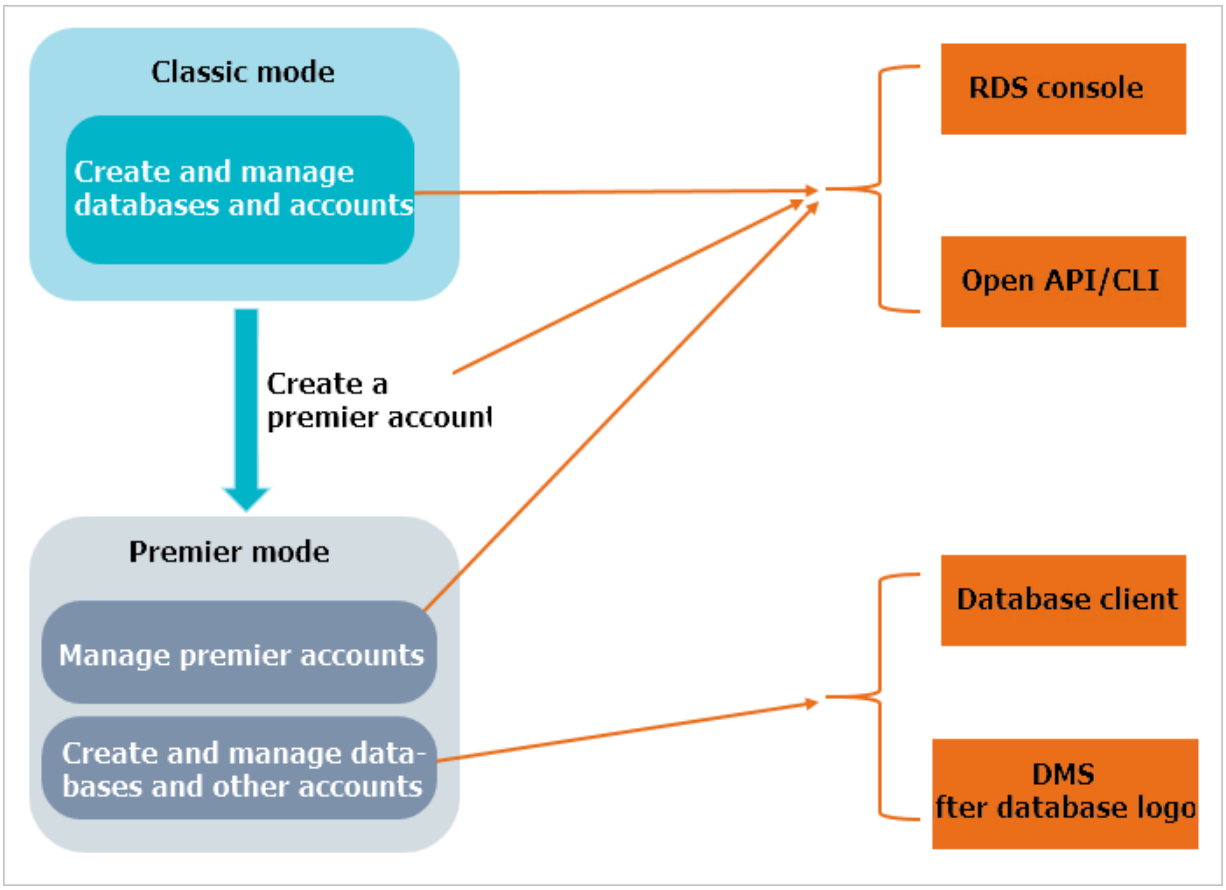
RDS supports two account management modes: classic and privileged. For ApsaraDB RDS for MySQL 5.6 or 5.7 instances, you can upgrade the account management mode from classic to privileged by creating a privileged account.

Context

Compared with the classic mode, the privileged mode enables more permissions and allows you to use SQL to directly manage databases and accounts. Therefore, we recommend that you use the privileged mode. After a privileged account is created for a primary instance, the privileged account is synchronized to read-only instances.

For ApsaraDB RDS for MySQL 5.6 or 5.7 instances, *Figure 9-2: Comparison of account management modes* **shows the differences between the two modes in creating and managing databases and accounts.**

Figure 9-2: Comparison of account management modes



Note:

- **In an ApsaraDB RDS for MySQL 5.6 or 5.7 instance, the account management mode can be upgraded from classic to privileged, but cannot be downgraded from privileged to classic.**
- **The instance restarts when a privileged account is created, causing a transient disconnection of less than 30 seconds. To avoid service impacts from transient disconnections, choose an appropriate time and ensure that your applications can be automatically reconnected.**

- **The following changes occur after an instance is upgraded to the privileged mode:**
  - **You cannot use the RDS console or API operations to manage databases and standard accounts. Instead, you must use SQL commands or DMS. The Create Account button is not displayed on the Accounts page.**
  - **In MySQL 5.6, you cannot directly access the mysql.user or mysql.db tables . However, you can view existing accounts and permissions through mysql. user_view and mysql.db_view.**
  - **In MySQL 5.6, you cannot use the privileged account to change the passwords of standard accounts. To change the password of a standard account, you must delete the account and create a new one.**
  - **You can use the RDS console or API operations to reset the password and permissions of a privileged account. The resetting operation does not affect the other accounts that have been created.**

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane of the Basic Information page, choose Database Management > Accounts.**

4. **On the Accounts page, click Create Account. On the Create Account page that appears, configure the parameters as prompted.**

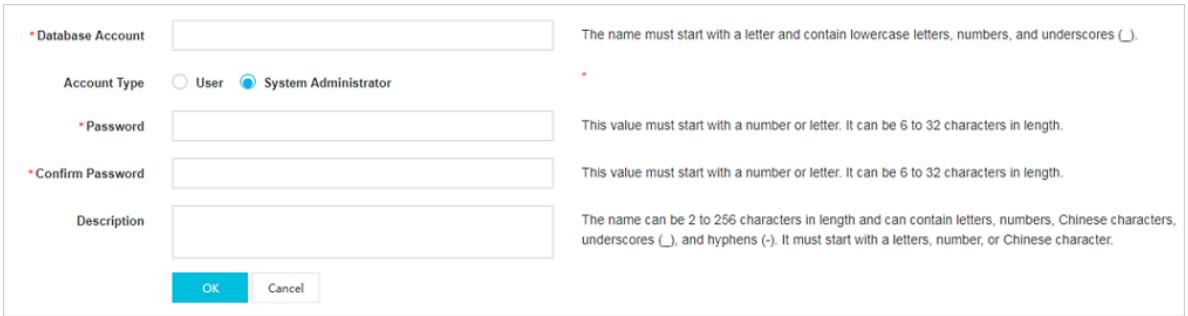
   Privileged account creation parameters **describes the parameter configurations.**

   Table 9-6: Privileged account creation parameters

Parameter	Description
Database Account	**The name of the account. The naming conventions are as follows:** **· It must start with a lowercase letter and end with a lowercase letter or digit.** **· It can contain lowercase letters, digits, and underscores (_).** **· It must be 2 to 16 characters in length.** **· It cannot contain reserved keywords. For more information about the reserved keywords, see the relevant document in the official website of the corresponding engine.**
Account Type	**The type of the account. The following types are available:** · `User` · `Super Administrator`**: This option is selected here.**
Password	**The password for the account. The requirements are as follows:** **· It must start with a digit or letter.** **· It can contain digits, letters, and underscores (_).** **· It must be 6 to 32 characters in length.**
Confirm Password	**Enter the password again.**

Parameter	Description
Description	The description of the account. The requirements are as follows: · It must start with a letter or digit. · It can contain letters, digits, underscores (_), and hyphens (-). · It must be 2 to 256 characters in length.

Figure 9-3: Create an account



5. After you set the parameters, click OK.

9.3.4.1.3 Create a standard account

After you create an RDS instance and configure the whitelist, you must create a database and an account in the instance. This topic describes how to create a standard account.

Context

To migrate the local database to ApsaraDB for RDS, you must create a database and an account in the RDS instance identical to those in the local database. Databases in an instance share all resources that belong to the instance.

Use service roles to create accounts and follow the principle of least privilege to assign appropriate read-only and read/write permissions to accounts. When necessary, you can create multiple database accounts and allow each of them to access data only relevant to their own business tasks.

Notice:
To ensure database security, set strong account passwords and change the passwords on a regular basis.

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane of the Basic Information page, choose Database Management > Accounts.**
4. **On the Accounts page, click Create Account. On the Create Account page that appears, configure the parameters as prompted.**

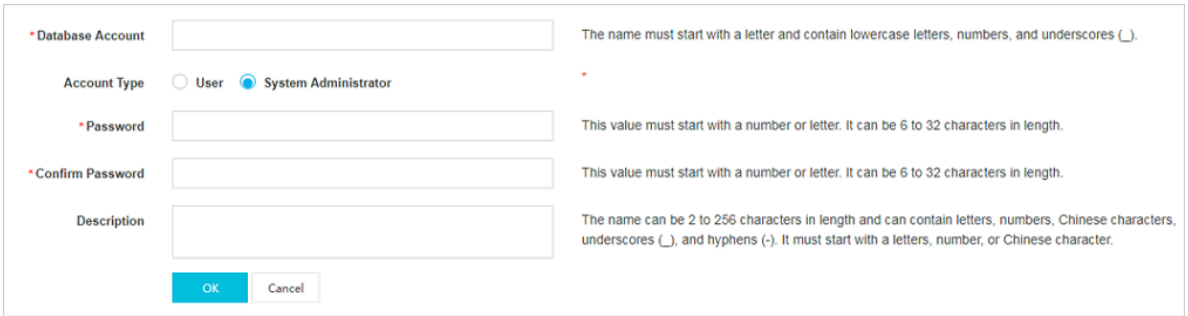
   Table 9-7: Standard account creation parameters **describes the parameter configurations.**

   Table 9-7: Standard account creation parameters

Parameter	Description
Database Account	The name of the account. The naming conventions are as follows: · It must start with a lowercase letter and end with a lowercase letter or digit. · It can contain lowercase letters, digits, and underscores (_). · It must be 2 to 16 characters in length. · It cannot contain reserved keywords. For more information about the reserved keywords, see the relevant document in the official website of the corresponding engine.
Account Type	The type of the account. The following types are available: · `User`: This option is selected here. · `Super Administrator`.
Password	The password for the account. The requirements are as follows: · It must start with a digit or letter. · It can contain digits, letters, and underscores (_). · It must be 6 to 32 characters in length.
Confirm Password	Enter the password again.

Parameter	Description
Description	The description of the account. The requirements are as follows: · It must start with a letter or digit. · It can contain letters, digits, underscores (_), and hyphens (-). · It must be 2 to 256 characters in length.

Figure 9-4: Create an account



5. After you set the parameters, click OK.

9.3.4.1.4 Create a database

After you create an RDS instance and configure the whitelist, you must create a database and an account in the instance.

Context

To migrate the local database to ApsaraDB for RDS, you must create a database and an account in the RDS instance identical to those in the local database. Use service roles to create accounts and follow the principle of least privilege to assign appropriate read-only and read/write permissions to accounts. When necessary, you can create multiple database accounts and allow each of them to access data only relevant to their own business tasks.

Note:
In the privileged mode, you are not permitted to manage databases and standard accounts in the RDS console or through API operations. You must use SQL statements or DMS to perform related operations.

Procedure

1. *Log on to the RDS console*.

2. **Click the ID of an instance.**
3. **In the left-side navigation pane of the Basic Information page, choose Database Management > Databases.**
4. **On the Databases page, click Create Database. On the Create Database page that appears, configure the parameters as prompted.**

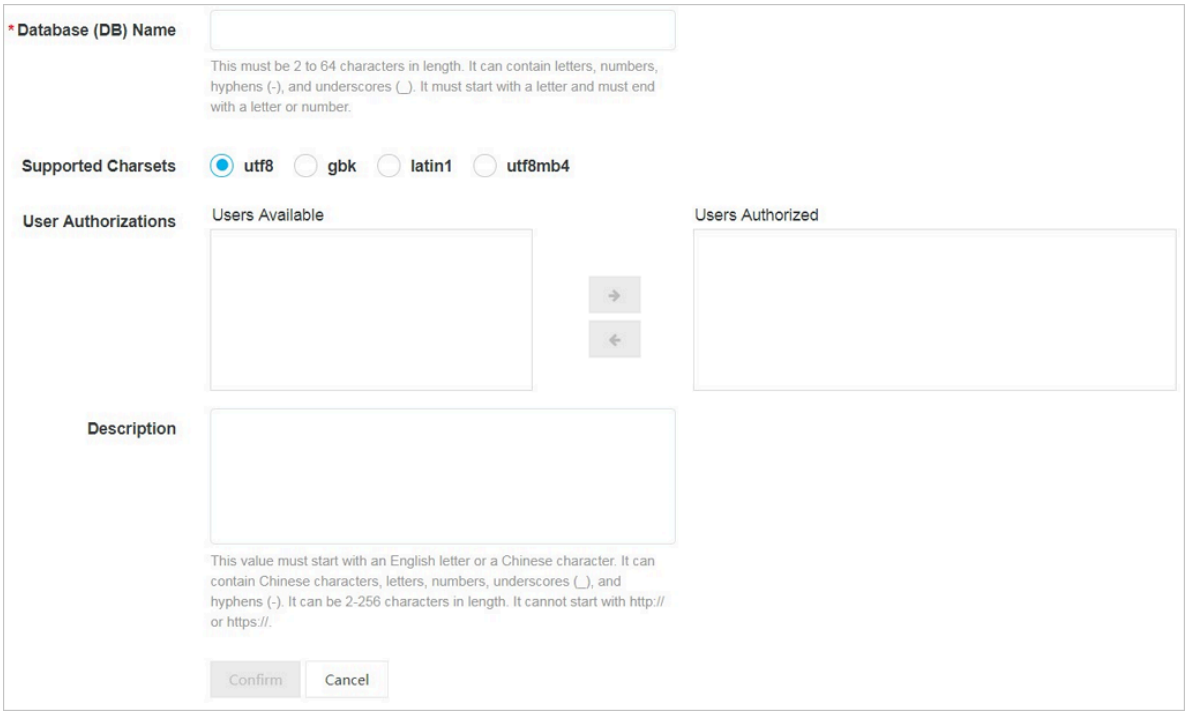
   Table 9-8: Database creation parameters **describes the parameter configurations.**

   Table 9-8: Database creation parameters

Parameter	Description
Database (DB) Name	The database name. The naming conventions are as follows: · It must start with a lowercase letter and end with a lowercase letter or digit. · It can contain lowercase letters, digits, underscores (_), and hyphens (-). · It must be 2 to 64 characters in length. · It cannot contain reserved keywords. For more information about the reserved keywords, see the relevant document in the official website of the corresponding engine.
Supported Character Set	The following character sets are supported by the database: · `utf-8` · `gbk` · `latin1` · `utf8mb4`
User Authorizations	Select an account to grant read-only or read/write permissions. **Note:** · By default, the privileged account is authorized to use the database. · The permissions on the database can be granted only to standard accounts.

Parameter	Description
Description	The description of the database, which facilitates database management. The requirements are as follows: · It must start with a lowercase letter. · It can contain lowercase letters, digits, underscores (_), and hyphens (-). · It must be 2 to 256 characters in length. · It cannot start with http:// or https://.

Figure 9-5: Create a database



5. After you set the parameters, click OK.

9.3.4.2 RDS for PostgreSQL

9.3.4.2.1 Configure a whitelist

To guarantee database security and reliability, you must modify the whitelist of an RDS instance before you enable the instance. You must add the IP addresses or CIDR blocks that are used for database access to the whitelist of the RDS instance.

Context

The system creates a default whitelist group for each instance. This whitelist group can be modified or cleared, but cannot be deleted. For each newly created RDS

instance, IP address 0.0.0.0/0 is added to the default whitelist group by default. 0.0.0.0/0 allows all IP addresses to access the instance, which greatly reduces database security. Delete 0.0.0.0/0 from the whitelist.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane of the Basic Information page, choose Security Control > Whitelist Settings.
4. You can use the following methods to add an IP address or CIDR block:
   - Add an IP address or CIDR block to the default whitelist group.

   a) Click the icon corresponding to the default whitelist group, and add an IP address or CIDR block.

   b) Click OK.
   - Create a whitelist group and add an IP address or CIDR block to the group.

   a) Click Create Whitelist Group. In the Create Whitelist Group dialog box that appears, set Group Name and IP Addresses.

   b) Click OK.

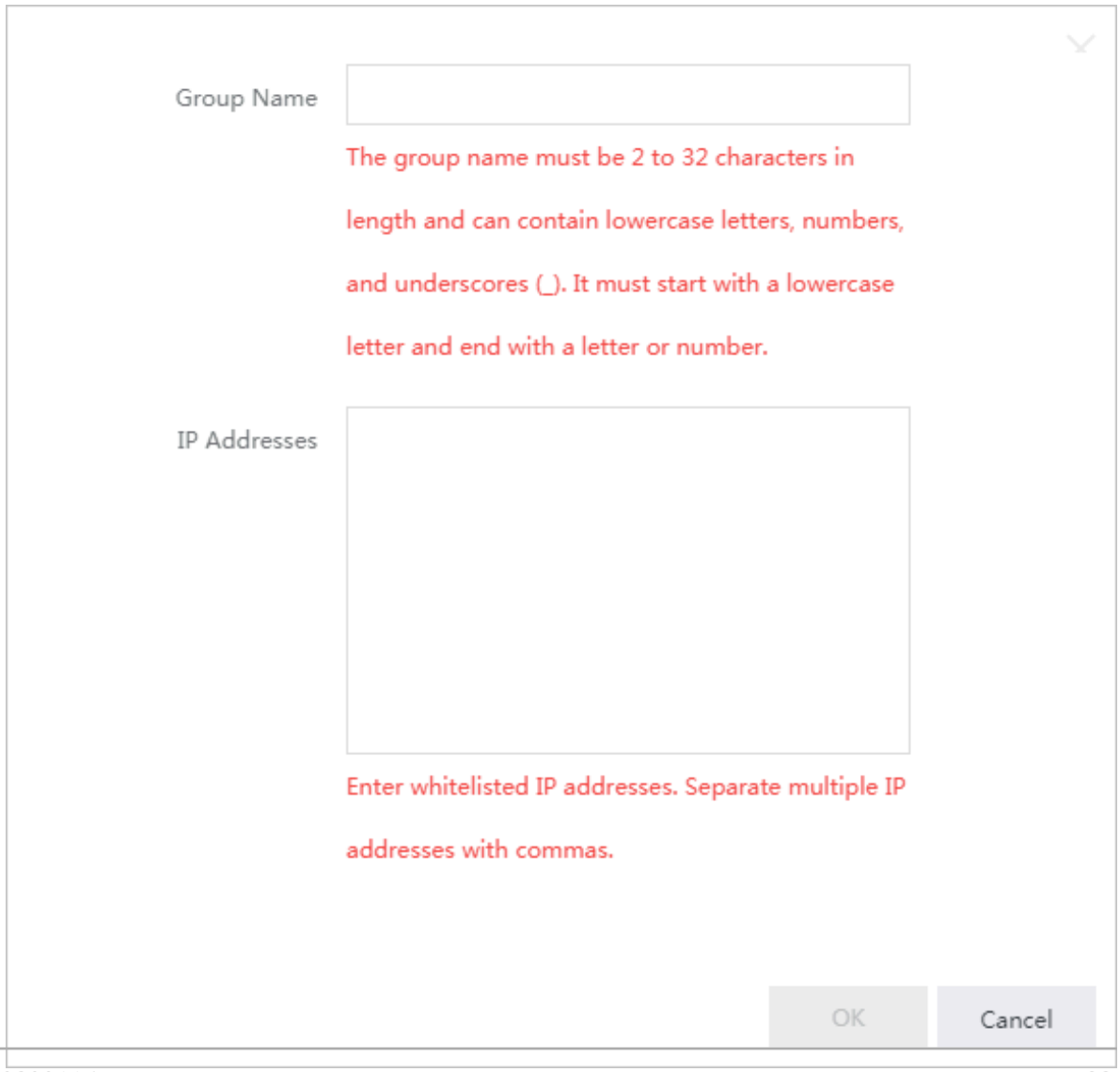
Table 9-9: Whitelist configuration parameters describes the parameter configurations.

Table 9-9: Whitelist configuration parameters

Parameter	Description
Group Name	The name of the new whitelist group. The naming conventions are as follows: · It must start with a lowercase letter and end with a lowercase letter or digit. · It can contain lowercase letters, digits, and underscores (_). · It must be 2 to 32 characters in length. Note: You cannot change the name of an existing whitelist group.

Parameter	Description
IP Addresses	The IP addresses or CIDR blocks that are allowed to access the RDS instance. **Note:** · If you enter a CIDR block, such as 10.10.10.0/24, any IP addresses in the 10.10.10.X format can access the RDS instance. · If you enter multiple IP addresses, use commas (,) to separate them. Example: 192.168.0.1,172.16.213.9 · 127.0.0.1 indicates that no IP addresses are allowed to access the RDS instance. · 0.0.0.0/0 indicates that all IP addresses are allowed to access the RDS instance.

Figure 9-6: Create a whitelist group



Notice:

The proper use of the whitelist can improve the access security of your RDS instance. We recommend that you maintain the whitelist on a regular basis. After you configure the whitelist, you can perform the following operations:

- **Click the icon to modify the whitelist group.**
- **Click the icon to clear the default whitelist group or delete a custom whitelist group.**

9.3.4.2.2 Create accounts and databases

After you create an RDS instance and configure the whitelist, you must create an initial account and database before you can connect to the instance.

Context

To migrate an on-premises database to ApsaraDB for RDS, you must create a database with the same name and an account with the same name and password in the RDS instance. Follow the least privilege principle to create accounts and assign role permissions. In addition, assign appropriate read-only or read/write permissions to accounts. When necessary, you can create multiple database accounts and allow each of them to access data only relevant to their own business tasks.

Notice:

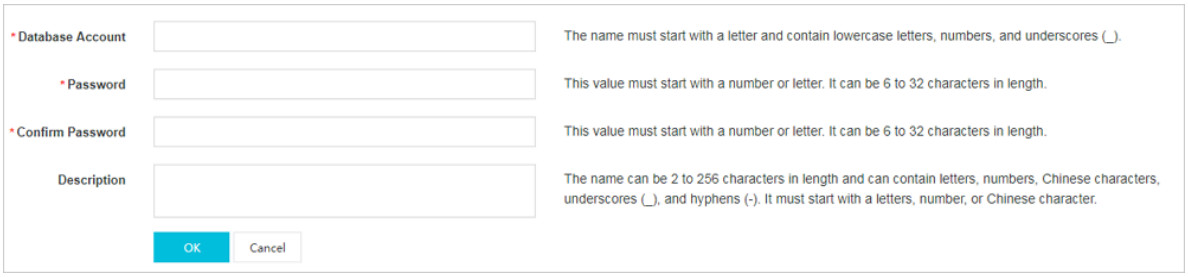
To ensure database security, set strong account passwords and change the passwords on a regular basis.

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane of the Basic Information page, choose Database Management > Accounts.**

4. On the Accounts page, click Create Account. On the Create Account page that appears, configure the following parameters.

Parameter	Description
Database Account	The name of the account. The naming conventions are as follows: · It must start with a lowercase letter and end with a lowercase letter or digit. · It can contain lowercase letters, digits, and underscores (_). · It must be 2 to 16 characters in length. · It cannot contain reserved keywords. For more information about the reserved keywords, see the relevant document in the official website of the corresponding engine.
Password	The password for the account. The requirements are as follows: · It must start with a lowercase letter and end with a lowercase letter or digit. · It can contain lowercase letters, digits, and underscores (_). · It must be 6 to 32 characters in length.
Confirm Password	Enter the password again.
Description	The description of the account. The requirements are as follows: · It must start with a letter or digit. · It can contain letters, digits, underscores (_), and hyphens (-). · It must be 2 to 256 characters in length.



5. Click OK.

> **Note:**
> After the initial account is created, you cannot use the console to delete the account or add other accounts. However, you can connect to the instance and execute SQL statements to create other accounts.

6. Connect to your RDS instance from the client. For more information, see *Connect to a PostgreSQL instance from a client*.
7. Execute the following statement to create a database:

   ```
   CREATE DATABASE"databasename"
   ```

   databasename specifies the name of the database to be created. For example, you can execute `CREATE DATABASE "mydatabase"`.

9.3.4.3 RDS for PPAS

9.3.4.3.1 Configure a whitelist

To guarantee database security and reliability, you must modify the whitelist of an RDS instance before you enable the instance. You must add the IP addresses or CIDR blocks that are used for database access to the whitelist of the RDS instance.

Context

The system creates a default whitelist group for each instance. This whitelist group can be modified or cleared, but cannot be deleted. For each newly created RDS instance, IP address 0.0.0.0/0 is added to the default whitelist group by default. 0.0.0.0/0 allows all IP addresses to access the instance, which greatly reduces database security. Delete 0.0.0.0/0 from the whitelist.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane of the Basic Information page, choose Security Control > Whitelist Settings.

4. You can use the following methods to add an IP address or CIDR block:

   - Add an IP address or CIDR block to the default whitelist group.

   a) Click the  icon corresponding to the default whitelist group, and add an IP address or CIDR block.

   b) Click OK.

   - Create a whitelist group and add an IP address or CIDR block to the group.

   a) Click Create Whitelist Group. In the Create Whitelist Group dialog box that appears, set Group Name and IP Addresses.

   b) Click OK.

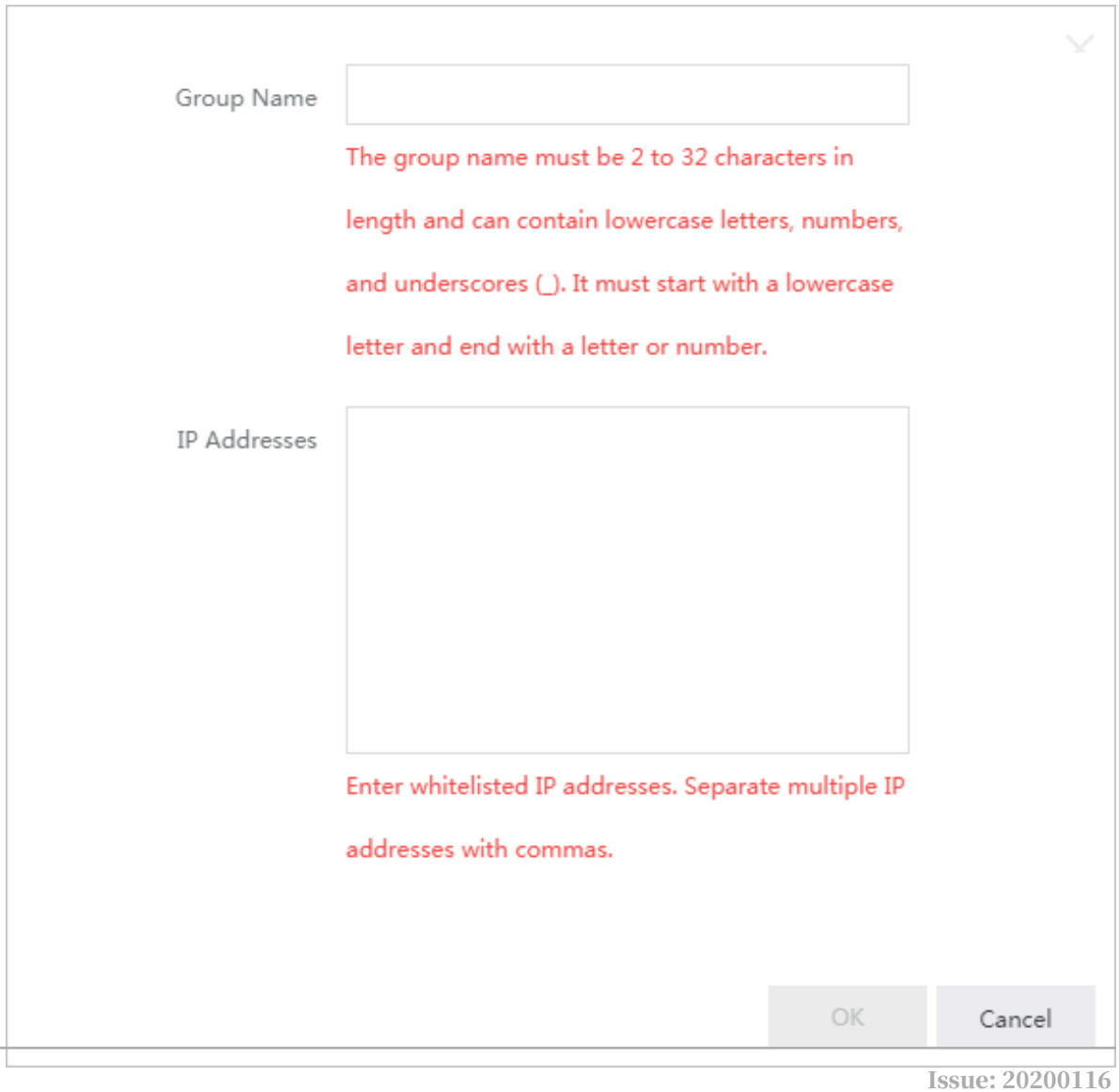
   Table 9-10: Whitelist configuration parameters describes the parameter configurations.

Table 9-10: Whitelist configuration parameters

Parameter	Description
Group Name	The name of the new whitelist group. The naming conventions are as follows: · It must start with a lowercase letter and end with a lowercase letter or digit. · It can contain lowercase letters, digits, and underscores (_). · It must be 2 to 32 characters in length. **Note:** You cannot change the name of an existing whitelist group.

Parameter	Description
IP Addresses	The IP addresses or CIDR blocks that are allowed to access the RDS instance. **Note:** · If you enter a CIDR block, such as 10.10.10.0/24, any IP addresses in the 10.10.10.X format can access the RDS instance. · If you enter multiple IP addresses, use commas (,) to separate them. Example: 192.168.0.1,172.16.213.9 · 127.0.0.1 indicates that no IP addresses are allowed to access the RDS instance. · 0.0.0.0/0 indicates that all IP addresses are allowed to access the RDS instance.

Figure 9-7: Create a whitelist group



Notice:

The proper use of the whitelist can improve the access security of your RDS instance. We recommend that you maintain the whitelist on a regular basis. After you configure the whitelist, you can perform the following operations:

- **Click the icon to modify the whitelist group.**
- **Click the icon to clear the default whitelist group or delete a custom whitelist group.**

9.3.4.3.2 Create accounts and databases

After you create an RDS instance and configure the whitelist, you must create a database and an account in the instance.

Context

To migrate the local database to ApsaraDB for RDS, you must create a database and an account in the RDS instance identical to those in the local database. Use service roles to create accounts and follow the principle of least privilege to assign appropriate read-only and read/write permissions to accounts. When necessary, you can create multiple database accounts and allow each of them to access data only relevant to their own business tasks.

Notice:

To ensure database security, set strong account passwords and change the passwords on a regular basis.

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane of the Basic Information page, choose Database Management > Accounts.**

4. **On the Accounts page, click Create Account. On the Create Account page that appears, configure the parameters as prompted.**

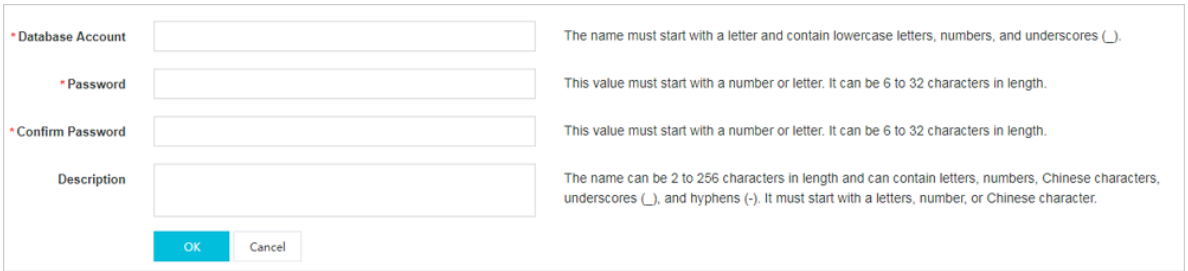
   Table 9-11: Account creation parameters **describes the parameter configurations.**

   Table 9-11: Account creation parameters

Parameter	Description
Database Account	The name of the account. The naming conventions are as follows: · It must start with a lowercase letter and end with a lowercase letter or digit. · It can contain lowercase letters, digits, and underscores (_). · It must be 2 to 16 characters in length. · It cannot contain reserved keywords. For more information about the reserved keywords, see the relevant document in the official website of the corresponding engine.
Password	The password for the account. The requirements are as follows: · It must start with a lowercase letter and end with a lowercase letter or digit. · It can contain lowercase letters, digits, and underscores (_). · It must be 6 to 32 characters in length.
Confirm Password	Enter the password again.

Parameter	Description
Description	The description of the account. The requirements are as follows: · It must start with a letter or digit. · It can contain letters, digits, underscores (_), and hyphens (-). · It must be 2 to 256 characters in length.

Figure 9-8: Create an account



5. **After you set the parameters, click OK.**
6. **Connect to your RDS instance from the client. For more information, see** *Connect to a PPAS instance from a client*.
7. **Run the following command to create a database:**

   `CREATE DATABASE"databasename"`

   databasename specifies the name of the database to be created, such as `CREATE DATABASE "mydatabase"`.

9.3.5 Connect to an instance

9.3.5.1 Log on to an instance through DMS

This topic describes how to connect to an RDS instance through Data Management Service (DMS).

Context

In the RDS console, you can log on to an RDS instance through DMS. DMS offers an integrated solution to view BI charts and data trends, track data, optimize performance, implement access security, and manage data, schemas, and servers. It can manage relational databases such as MySQL, PostgreSQL, and DRDS as well as OLAP databases such as AnalyticDB.

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of an instance.**
3. **On the Basic Information page, click Log On to DMS. On the logon page of the DMS console, enter the correct logon information as prompted.**

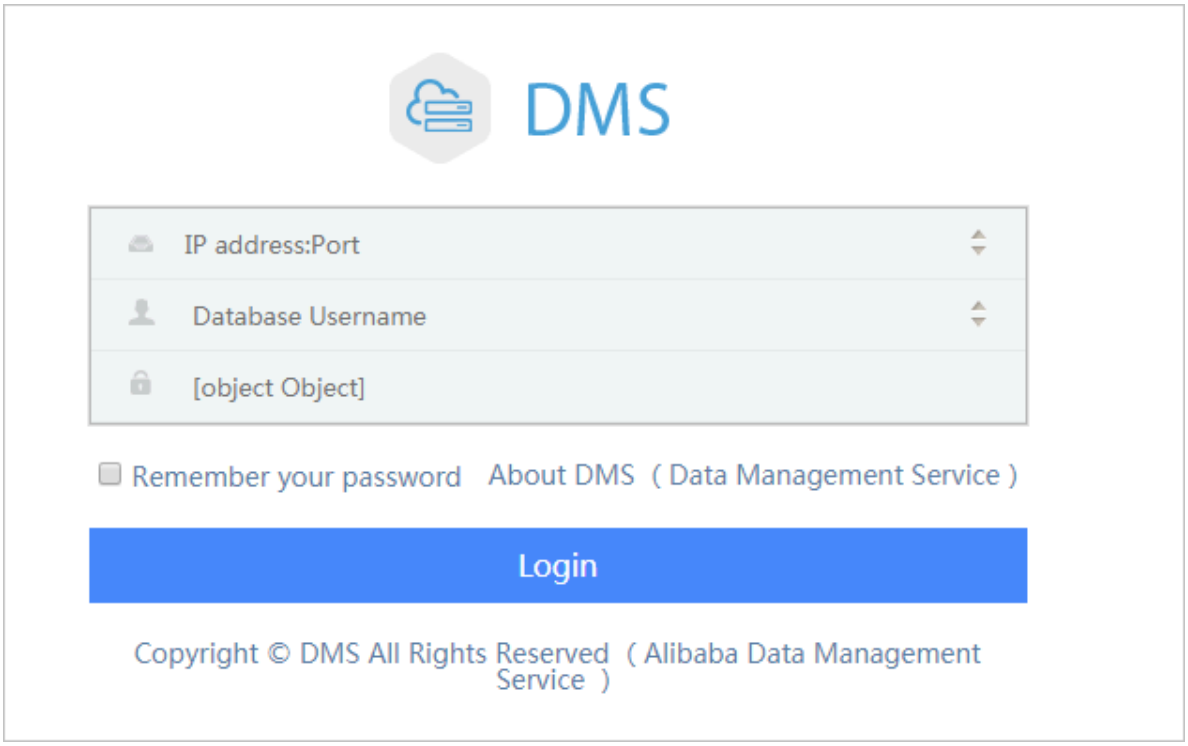
   Table 9-12: DMS logon parameters **describes the logon information.**

   Table 9-12: DMS logon parameters

Parameter	Description
IP address: Port	**The internal IP address and port number that are used to connect to the instance, such as rm-test000001.mysql.aliyun-inc.com:3306.** **To obtain the internal IP address and port number, perform the following steps:** **a.** *Log on to the RDS console*. **b. Click the ID of the instance to go to the Basic Information page.** **c. In the Internal Network Connection Information section, view the internal IP address and port number of the instance.**
Database Username	**The account that is used to connect to the RDS instance. It is the account that you created in the instance. For more information about how to create an account in a MySQL instance, see** *Create a standard account* **or** *Create a privileged account*.

Parameter	Description
Password	The password for the account that is used to connect to the RDS instance. It is the password that you specified for the account created in the instance.

Figure 9-9: DMS logon page



4. After you set the parameters, click Logon.

Note:
If you want the Web browser to remember the password, select `Remember your password` and click Logon.

9.3.5.2 Connect to a MySQL instance from a client

This topic describes how to connect to an ApsaraDB RDS for MySQL instance from the MySQL-Front client.

Prerequisites

- You have installed the MySQL-Front client.
- Your client is deployed in the same VPC as the RDS instance.

- You have added the IP address used to access the RDS instance to the RDS whitelist. For more information about how to configure a whitelist, see *Configure a whitelist*.

Context

ApsaraDB RDS for MySQL is fully compatible with the native database service. You can connect to RDS in the same way you connect to an on-premises MySQL server. You can refer to this topic as an example when you connect from other clients.

Procedure

1. Start the MySQL-Front client on your PC.
2. In the Open Connection window, click New. In the Add Account dialog box that appears, configure the parameters as prompted.

   MySQL-Front logon parameters describes the parameter configurations.

   Table 9-13: MySQL-Front logon parameters

Parameter	Description
Name	The name of the database connection task. If this parameter is left blank, the system will use the same value as provided in the Host field.
Host	The internal IP address that is used to connect to the RDS instance. To obtain the internal IP address and port number, perform the following steps: a. *Log on to the RDS console*. b. Click the ID of the instance to go to the Basic Information page. c. In the Internal Network Connection Information section, view the internal IP address and port number of the instance.
Port	The internal network port that is used to connect to the RDS instance.
User	The account that is used to connect to the RDS instance. It is the account that you created in the instance.

Parameter	Description
Password	The password for the account that is used to connect to the RDS instance. It is the password that you specified for the account created in the instance.

Figure 9-10: Create a connection

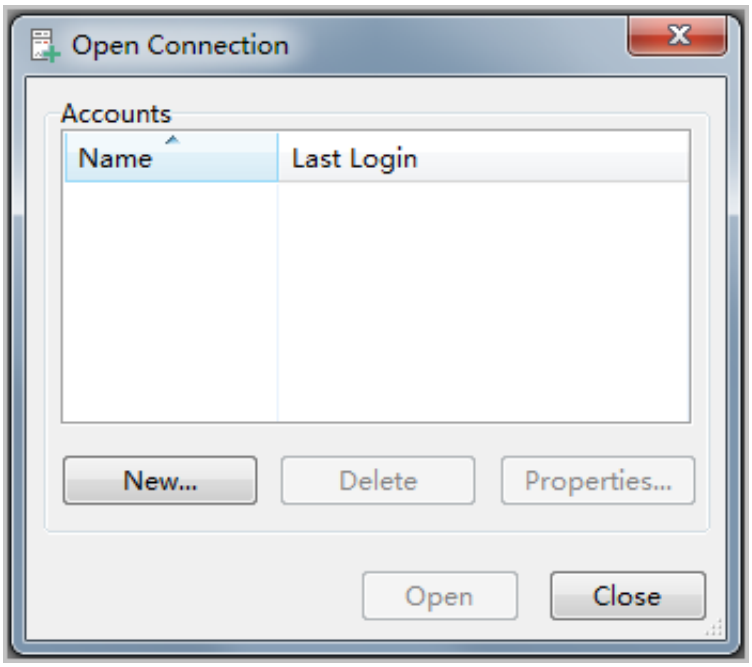
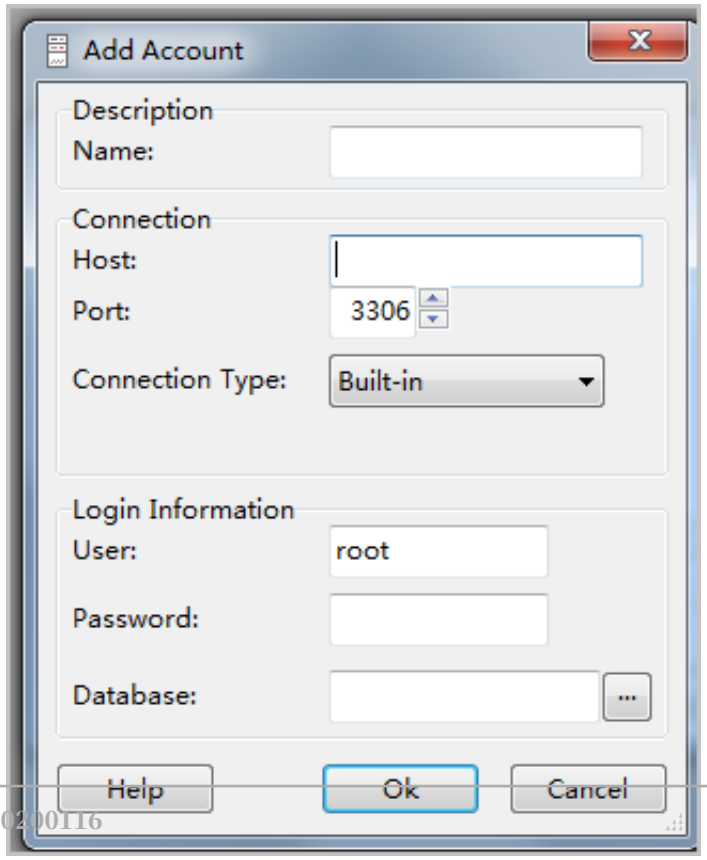


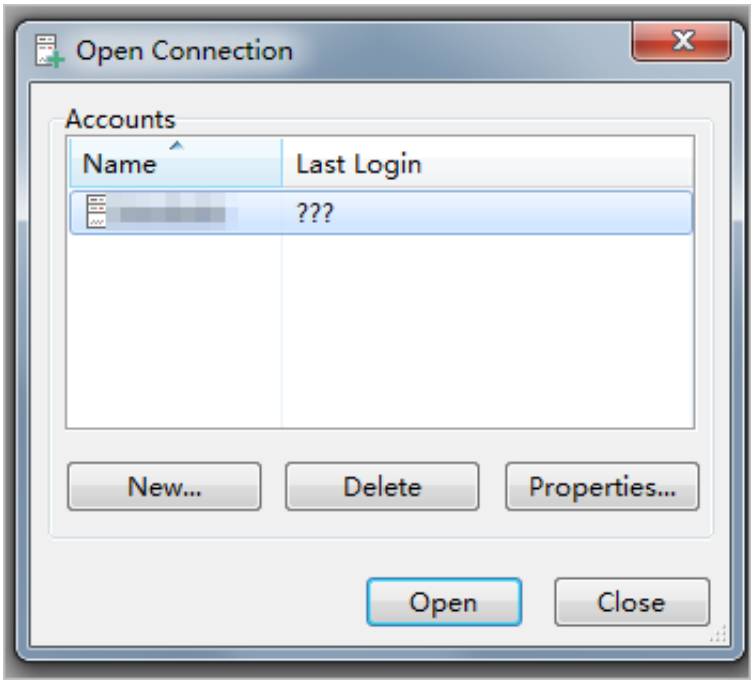
Figure 9-11: Enter the connection information



3. **After you set the parameters, click OK.**
4. **In the Open Connection window, select the created connection and click Open, as shown in** *Figure 9-12: Connect to an instance*.

> **Note:**
> **If the connection information is correct, the connection to the RDS instance will succeed.**

Figure 9-12: Connect to an instance



9.3.5.3 Connect to a PostgreSQL instance from a client

This topic describes how to connect to an ApsaraDB RDS for PostgreSQL instance from a pgAdmin 4 client.

Prerequisites

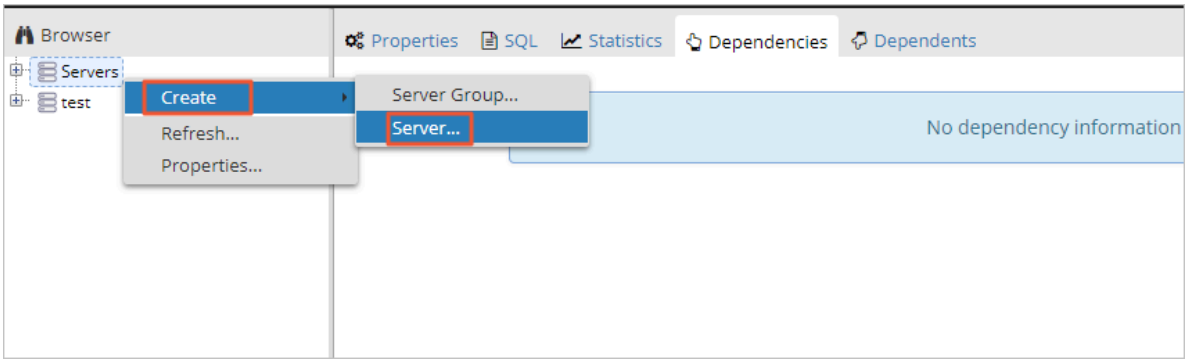
- **You have installed the pgAdmin 4 client.**
- **Your client is deployed in the same VPC as the RDS instance.**
- **You have added the IP address used to access the RDS instance to the RDS whitelist. For more information about how to configure a whitelist, see** *Configure a whitelist*.

Procedure

1. **Start the pgAdmin 4 client on your PC.**

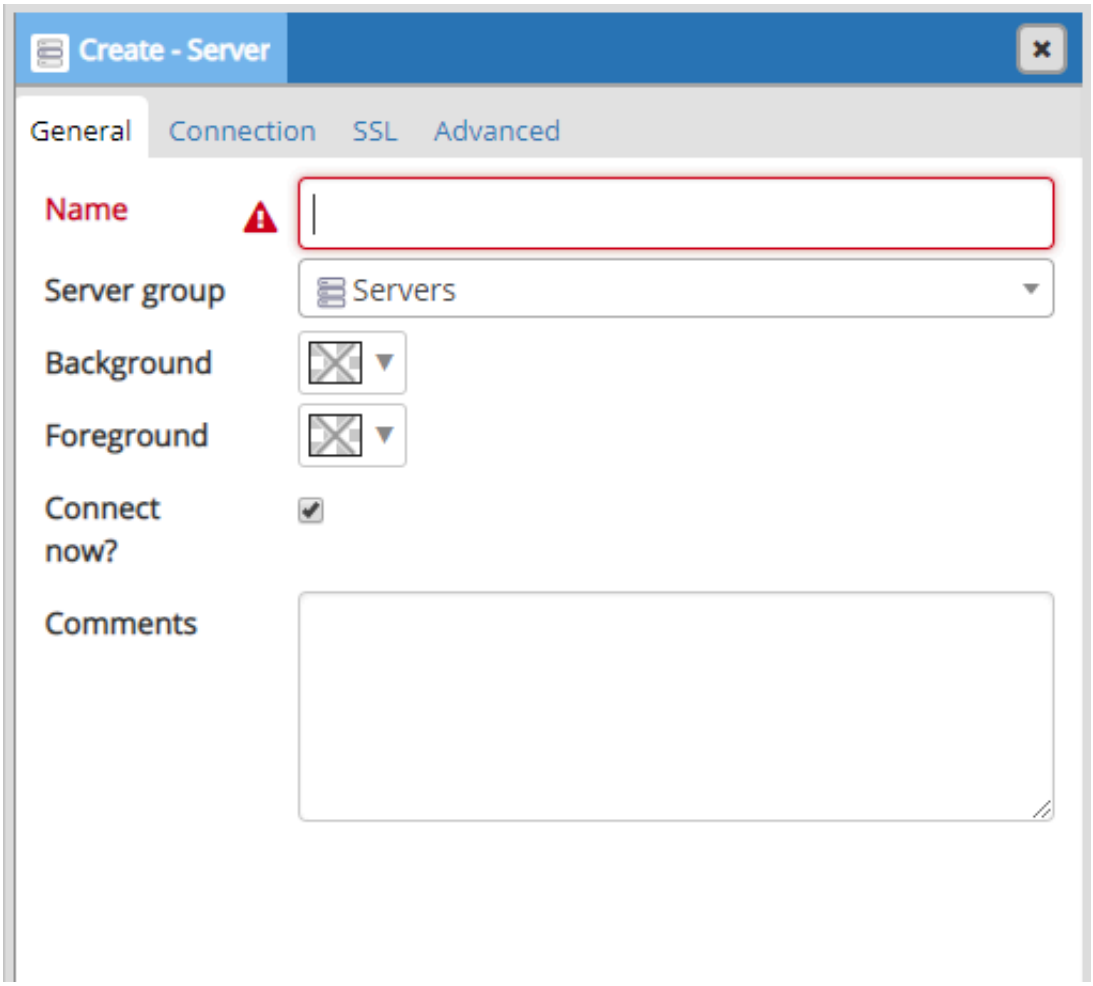
2. **Right-click Servers and choose Create > Server from the shortcut menu.**

Figure 9-13: Create a server



3. **On the Create - Server page, click the General tab and configure the parameters as prompted.**
   - **Name: specifies the name of the server to be created. Select an appropriate name so that the server can be easily located.**
   - **Comments: specifies additional information about the server to be created.**

Figure 9-14: Enter a server name



4. **Click the Connection tab and configure the parameters as prompted.**

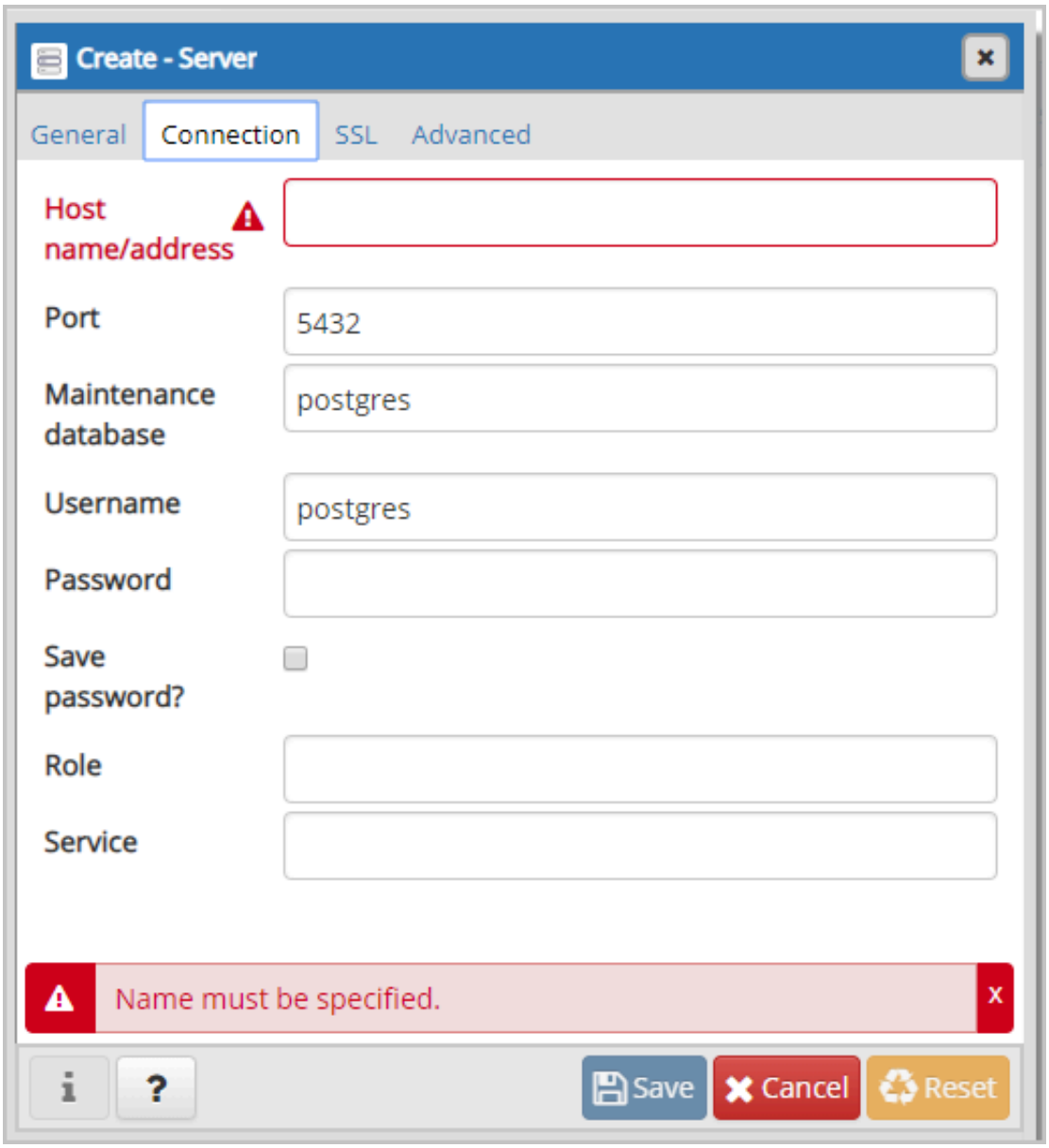
   Table 9-14: pgAdmin 4 logon parameters **describes the parameter configurations.**

   Table 9-14: pgAdmin 4 logon parameters

Parameter	Description
Host name/ address	The internal IP address that is used to connect to the RDS instance. To obtain the internal IP address and port number, perform the following steps: a. *Log on to the RDS console*. b. Click the ID of the instance to go to the Basic Information page. c. In the Internal Network Connection Information section, view the internal IP address and port number of the instance.
Port	The internal network port that is used to connect to the RDS instance.
Username	The account that is used to connect to the RDS instance. It is the account that you created in the instance.

Parameter	Description
Password	The password for the account that is used to connect to the RDS instance. It is the password that you specified for the account created in the instance.

Figure 9-15: Configure the instance connection information



5. After you set the parameters, click Save.
6. If the connection information is correct, choose Servers > server name > Databases > postgres. The server name parameter varies in actual scenarios.

9.3.5.4 Connect to a PPAS instance from a client

This topic describes how to connect to an ApsaraDB RDS for PPAS instance from a pgAdmin 4 client.

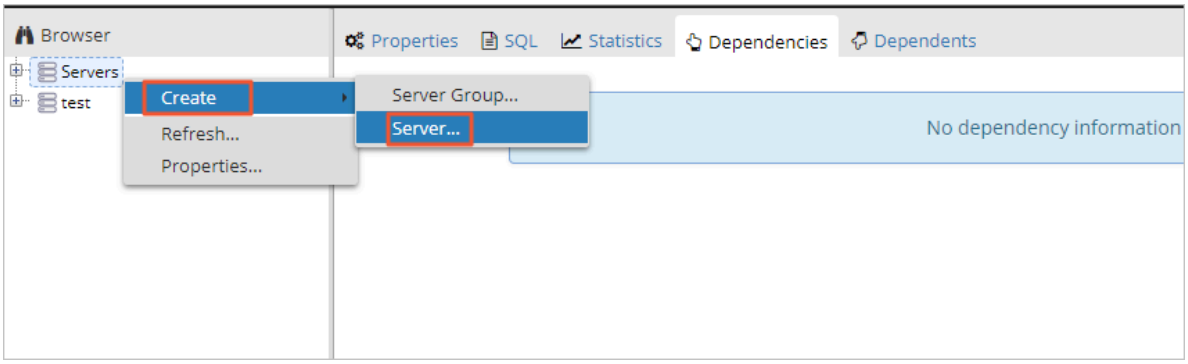
Prerequisites

- **You have installed the pgAdmin 4 client.**
- **Your client is deployed in the same VPC as the RDS instance.**
- **You have added the IP address used to access the RDS instance to the RDS whitelist. For more information about how to configure a whitelist, see** *Configure a whitelist*.

Procedure

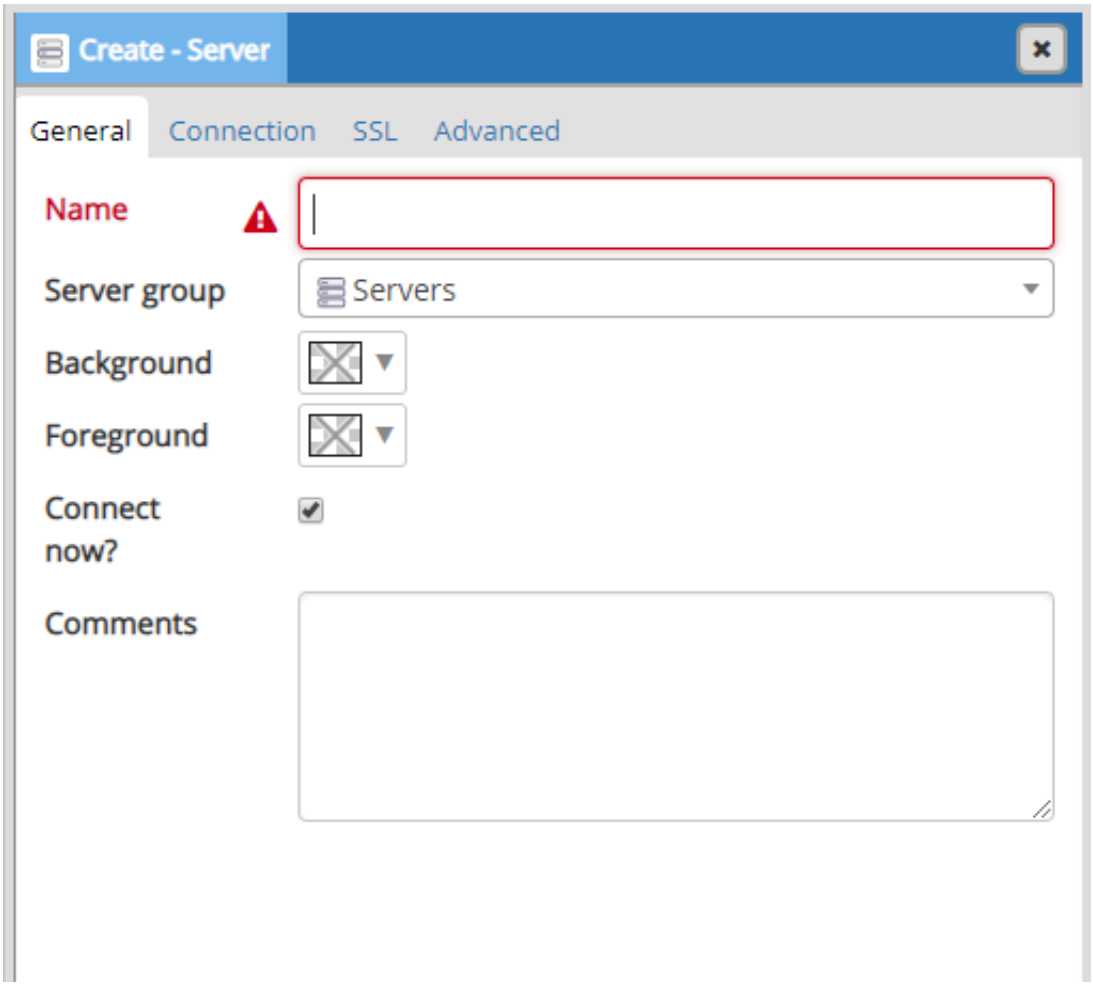
1. **Start the pgAdmin 4 client on your PC.**
2. **Right-click Servers and choose Create > Server from the shortcut menu.**

Figure 9-16: Create a server



3. **On the Create - Server page, click the General tab and configure the parameters as prompted.**
   - **Name: specifies the name of the server to be created. Select an appropriate name so that the server can be easily located.**
   - **Comments: specifies additional information about the server to be created.**

Figure 9-17: Enter a server name



4. **Click the Connection tab and configure the parameters as prompted.**

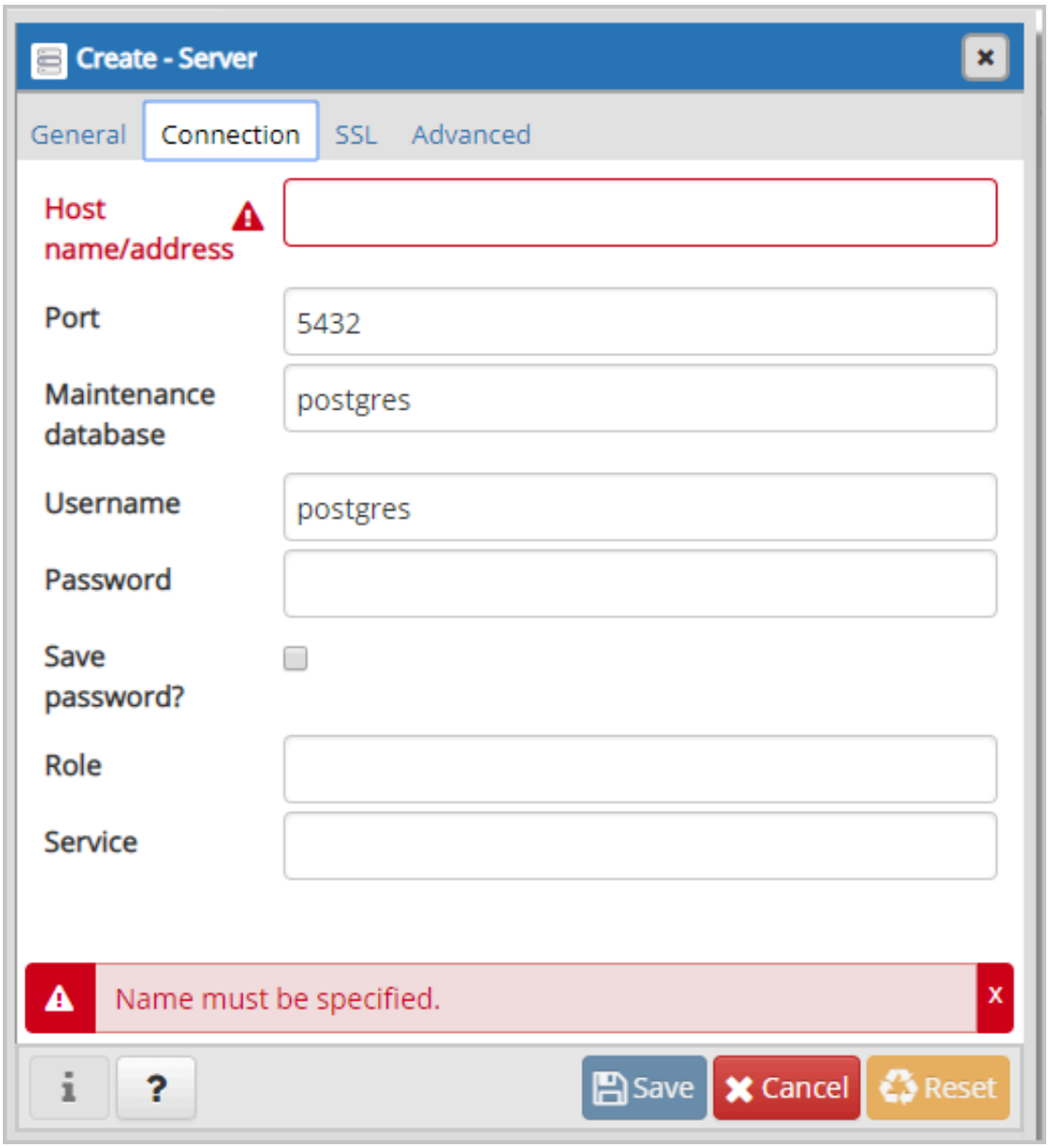
   Table 9-15: pgAdmin 4 logon parameters **describes the parameter configurations.**

   Table 9-15: pgAdmin 4 logon parameters

Parameter	Description
Host name/ address	The internal IP address that is used to connect to the RDS instance. To obtain the internal IP address and port number, perform the following steps: a. *Log on to the RDS console*. b. Click the ID of the instance to go to the Basic Information page. c. In the Internal Network Connection Information section, view the internal IP address and port number of the instance.
Port	The internal network port that is used to connect to the RDS instance.
Username	The account that is used to connect to the RDS instance. It is the account that you created in the instance.

Parameter	Description
Password	The password for the account that is used to connect to the RDS instance. It is the password that you specified for the account created in the instance.

Figure 9-18: Configure the instance connection information



5. After you set the parameters, click Save.
6. If the connection information is correct, choose Servers > server name > Databases > postgres. The server name parameter varies in actual scenarios.

9.4 Instances

9.4.1 Create an instance

This topic describes how to create an instance in the RDS console.

Prerequisites

Before you create an RDS instance, you must apply for an Apsara Stack tenant account.

Procedure

1. *Log on to the RDS console*.
2. **On the Relational Database Service (RDS) page, click Create Instance. On the Create Instance page that appears, configure the parameters as promoted.**

   The parameters are described as follows.

   Table 9-16: Instance creation parameters

Category	Parameter	Description
Basic Settings	Department	The department to which the instance belongs.
	Project	The project to which the instance belongs.
	Region	The region where the instance is located. Services in different regions are not interconnected over the internal network. Once a region is selected, it cannot be changed.
	Zone	The zone of the instance. Common RDS instances use a hot standby architecture. Single zone indicates that the primary and secondary nodes are in the same zone.
Specifications	Instance Name	The name of the RDS instance. It must start with a letter . It can contain letters, digits, underscores (_), and hyphens (-). It must be 2 to 64 characters in length.
	Database Type	The database type varies with region. The available database types are displayed on the Create Instance page.
	Database Version	The version of the database.

Category	Parameter	Description
	CPU/ Memory	The specifications of the instance. Available specifications are as follows: · Dedicated: This type of specifications is followed by the "Dedicated" suffix. · Dedicated Hosts: This type of specifications is followed by the "Dedicated Hosts" suffix. Memory size determines the maximum number of connections and IOPS. The actual values are displayed on the console UI.
	Storage Capacity	The storage space of the instance, including the space for data, system files, binlog files, and transaction files.
Network Type	Instance Type	The type of the instance. Based on your business scenario, you can select from either internal instance or external instance. External instances can only support the classic network type.
	Network Type	RDS instances support the following network types: · `Classic Network`: Cloud services in a classic network are not isolated. Unauthorized access to a cloud service is only blocked by the security group or whitelist policy of the service. · `VPC`: A VPC helps you to build an isolated network environment in Alibaba Cloud. You can customize the routing table, IP address range, and gateway in a VPC. We recommend that you select VPC for enhanced security. You can create a VPC in advance, or change the network type to VPC after creating an instance.

Category	Parameter	Description
Connection Mode	Connection Mode	RDS instances support two connection modes: `Standard Mode` and `Safe Mode`. · Standard Mode: RDS uses SLB to eliminate the impact of database engine HA switching on the application layer. This mode shortens the response time, but slightly increases the probability of transient disconnections and disables SQL interception. · Safe Mode: This mode prevents 90% of transient disconnections and intercepts SQL injection attacks based on semantic analysis. However, it increases the response time by over 20%.
Quantity	Instances	The number of RDS instances that can be created simultaneously. Maximum value: 20.

3. After you set the parameters, click Create.

9.4.2 View instance details

You can view the details of an instance, such as its basic information, intranet connection information, running status, and configurations.

Procedure

1. *Log on to the RDS console*.
2. You can use either of the following methods to view the instance details page:
   - Click the ID of the instance to go to the Basic Information page.
   - Click the icon in the Actions column corresponding to the instance and choose View Details to go to the Basic Information page.

9.4.3 Restart an instance

You can manually restart instances when the number of connections exceeds the threshold or performance issues occur on the instances.

Context

Note:
Restarting an instance will cause service interruptions. Exercise caution when performing this operation and make sure that the restart does not affect other services.

Procedure

1. *Log on to the RDS console*.
2. On the Basic Information page, click the icon in the Actions column corresponding to the instance that you want to restart and choose Restart Instance from the shortcut menu. In the Restart Instance message that appears, click OK.

9.4.4 Modify configurations

You can modify configurations of your instance, such as specifications and storage space, if the configurations do not meet the requirements of your application.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. On the Basic Information page, click Change Configuration.
4. On the Change Configuration page that appears, set Specifications and Storage Size (GB) for the instance.
5. Click OK.

9.4.5 Release an instance

You can manually release instances as needed.

Context

Note:

- You can only manually release instances in the running status.
- If read/write splitting is enabled for the primary instance, you must *Disable read/write splitting*.

Procedure

1. *Log on to the RDS console*.
2. Click the icon in the Actions column corresponding to the instance that you want to release, and choose Delete Instance from the shortcut menu.
3. In the Delete Instance message that appears, click OK.

9.4.6 Configure parameters

ApsaraDB for RDS allows you to define certain instance parameters. For more information about the modifiable parameters, see Parameter Settings in the RDS console.

Context

ApsaraDB for RDS is fully compatible with the native database service. The parameter configuration methods for both services are similar. This example uses the RDS console to modify the parameters. You can also call API operations and execute commands to modify the parameters.

Notice:

- **ApsaraDB RDS for PostgreSQL or PPAS instances do not support custom parameters.**
- **Configure parameters on the Parameter Settings page based on the specified Parameter Range.**
- **Modifying some parameters requires you to restart the instance. Go to the Parameter Settings page and check Requires Restart to determine whether a restart is required. Before you restart the instance, ensure that restarting the instance will not affect other services.**

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane of the Basic Information page, choose Performance Optimization > Parameter Settings.**
4. **On the Parameter Settings page, click the icon in the Actions column corresponding to the parameter to be configured and choose Change from the shortcut menu. In the Change Parameter dialog box that appears, configure the parameters as prompted.**

   Note:

   - **For more information about parameter settings, see *Modifiable MySQL instance parameters*.**

- **PostgreSQL or PPAS parameter oss_fdw.rds_oss_endpoint_whitelist is used to configure an OSS endpoint whitelist. You can use the oss_fdw plug-in to read and write external data.**

5. **After you set the parameters, click OK.**

9.4.7 Change ownership

You can change the department or project that an instance belongs to based on your business requirements.

Procedure

1. *Log on to the RDS console*.
2. **Click the icon in the Actions column corresponding to the instance that you want change and choose Change Ownership from the shortcut menu. In the Change Ownership dialog box that appears, configure the parameters as prompted.**

   Table 9-17: Ownership changing parameters **describes the parameter configurations.**

   Table 9-17: Ownership changing parameters

Parameter	Description
Instance Name	The name of the instance to which the ownership is to be transferred. The instance name is specified by the system and cannot be changed. For more information about how to change an instance name, see *Change the instance name*.
Department	The department to which the instance belongs.
Project	The project to which the instance belongs.

3. **Click OK.**

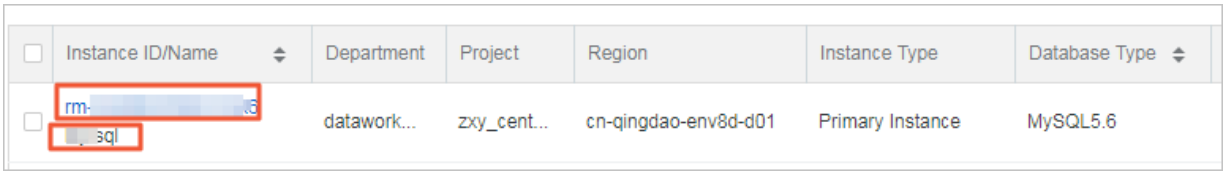
9.4.8 Change the instance name

You can change instance names for easy management.

Context

In the instance list, the`Instance ID/Name` column shows instance IDs in the upper part and instance names in the lower part, as shown in *Figure 9-19: Instance ID/Name*. You can change instance names but cannot change instance IDs.

Figure 9-19: Instance ID/Name



Procedure

1. *Log on to the RDS console*.
2. **Click the icon in the Actions column corresponding to the instance that you want to change and choose Change Instance Name from the shortcut menu. In the Change Instance Name dialog box that appears, change the instance name as prompted.**

   Note:

   - **The instance name must begin with a letter.**
   - **The instance name can contain letters, underscores (_), hyphens (-), and digits.**
   - **The instance name can be 2 to 64 characters in length.**
3. **Click OK.**

9.4.9 Change the port number

You can change the instance port number to facilitate management.

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of an instance.**
3. **On the Basic Information page, click Change Port Number.**
4. **Set a new port number.**

   Note:
   The port number must be in the range of 1000 to 65535.
5. **Click OK.**

9.4.10 Typical parameter settings

9.4.10.1 Modifiable MySQL instance parameters

This topic describes modifiable ApsaraDB RDS for MySQL parameters.

The following table lists the modifiable MySQL instance parameters. For more information about the parameters, see the MySQL official documentation at *https://dev.mysql.com/doc/***.**

Table 9-18: Modifiable MySQL 5.6 instance parameters

Parameter	Default value	Running parameter value	Restart required	Value range	Description
auto_increment_increment	1	1	No	[1-65535]	auto_increment_increment and auto_increment_offset are intended for use with master-to-master replication, and can be used to control the operation of AUTO_INCREMENT columns. Both variables have global and session values, and each can assume an integer value between 1 and 65,535 inclusive. Setting the value of either of these two variables to 0 causes its value to be set to 1 instead. Attempting to set the value of either of these two variables to an integer greater than 65,535 or less than 0 causes its value to be set to 65,535 instead. Attempting to set the value of auto_increment_increment or auto_increment_offset to a noninteger value produces an error, and the actual value of the variable remains unchanged.
auto_increment_offset	1	1	No	[1-65535]	determines the starting point for the AUTO_INCREMENT column value.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
back_log	3000	3000	Yes	[0-65535]	The number of outstanding connection requests that MySQL can have.
binlog_cache_size	128 KB	128 KB	No	[4096-16777216]	The size of the cache to hold changes to the binary log during a transaction.
binlog_checksum	CRC32	CRC32	Yes	[CRC32|NONE]	The master to write a checksum for each event in the binary log.
binlog_row_image	full	full	No	[full|minimal]	Binlog save every column or actually required column in binlog images.
binlog_stmt_cache_size	32768	32768	No	[4096-16777216]	The size of the statement cache for updates to non-transactional engines for the binary log.
block_encryption_mode	"aes-128-ecb"	"aes-128-ecb"	No	["aes-128-ecb"|" aes-192-ecb"|" aes-256-ecb"|" aes-128-cbc"|" aes-192-cbc"|" aes-256-cbc"]	This variable controls the block encryption mode for block-based algorithms such as AES. It affects encryption for AES_ENCRYPT() and AES_DECRYPT().
character_set_server	utf8	utf8	Yes	[utf8|latin1|gbk|utf8mb4]	The server's default character set.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
concurrent_insert	1	1	No	[0|1|2]	NEVER-Disables concurrent inserts; AUTO-(Default) Enables concurrent insert for MyISAM tables that do not have holes; ALWAYS -Enables concurrent inserts for all MyISAM tables, even those that have holes. For a table with a hole, new rows are inserted at the end of the table if it is in use by another thread . Otherwise, MySQL acquires a normal write lock and inserts the row into the hole.
connect_timeout	10	10	No	[1-3600]	The number of seconds that the mysqld server waits for a connect packet before responding with Bad handshake. The default value is 10 seconds as of MySQL 5.1.23 and 5 seconds before that.Increasing the connect_timeout value might help if clients frequently encounter errors of the form Lost connection to MySQL server at 'XXX', system error: errno.
default_storage_engine	InnoDB	TokuDB	Yes	[InnoDB| TokuDB|innodb |tokudb]	The default storage engine for new tables.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
default_time_zone	SYSTEM	SYSTEM	Yes	[SYSTEM|-12:00|-11:00|-10:00|-9:00|-8:00|-7:00|-6:00|-5:00|-4:00|-3:00|-2:00|-1:00|+0:00|+1:00|+2:00|+3:00|+4:00|+5:00|+5:30|+6:00|+6:30|+7:00|+8:00|+9:00|+10:00|+11:00|+12:00|+13:00]	The default time zone for the database.
default_week_format	0	0	No	[0-7]	The default mode value to use for the WEEK() function.
delayed_insert_limit	100	100	No	[1-4294967295]	After inserting delayed_insert_limit delayed rows, the INSERT DELAYED handler thread checks whether there are any SELECT statements pending. If so, it permits them to execute before continuing to insert delayed rows.
delayed_insert_timeout	300	300	No	[1-3600]	How many seconds an INSERT DELAYED handler thread should wait for INSERT statements before terminating.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
delayed_queue_size	1000	1000	No	[1-4294967295]	This is a per-table limit on the number of rows to queue when handling INSERT DELAYED statements. If the queue becomes full, any client that issues an INSERT DELAYED statement waits until there is room in the queue again.
delay_key_write	ON	ON	No	[ON|OFF|ALL]	This option applies only to MyISAM tables. It can have one of the following values to affect handling of the DELAY_KEY_WRITE table option that can be used in CREATE TABLE statements.
div_precision_increment	4	4	No	[0-30]	This variable indicates the number of digits by which to increase the scale of the result of division operations performed with the / operator. The default value is 4. The minimum and maximum values are 0 and 30, respectively. The following example illustrates the effect of increasing the default value.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
explicit_defaults_for_timestamp	false	false	Yes	true|false	As indicated by the warning, to turn off the nonstandard behaviors, enable the explicit_defaults_for_timestamp system variable at server startup.
ft_min_word_len	4	4	Yes	[1-3600]	The minimum length of the word to be included in a FULLTEXT index.
ft_query_expansion_limit	20	20	Yes	[0-1000]	The number of top matches to use for full-text searches performed using WITH QUERY EXPANSION.
group_concat_max_len	1024	1024	No	[4-18446744073709547520]	The maximum permitted result length in bytes for the GROUP_CONCAT() function. The default is 1024, Unit:Byte.
innodb_autoinc_lock_mode	1	1	Yes	[0|1|2]	The number of threads that can enter InnoDB concurrently is determined by the innodb_thread_concurrency variable.
innodb_concurrency_tickets	5000	5000	No	[1-4294967295]	The number of threads that can enter InnoDB concurrently is determined by the innodb_thread_concurrency variable.
innodb_ft_max_token_size	84	84	Yes	[10-84]	Maximum length of words that are stored in an InnoDB FULLTEXT index.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
innodb_ft_min_token_size	3	3	Yes	[0-16]	Minimum length of words that are stored in an InnoDB FULLTEXT index.
innodb_large_prefix	OFF	OFF	No	[ON|OFF]	Enable this option to allow index key prefixes longer than 767 bytes (up to 3072 bytes), for InnoDB tables that use the DYNAMIC and COMPRESSED row formats.
innodb_lock_wait_timeout	50	50	No	[1-1073741824]	The timeout in seconds an InnoDB transaction may wait for a row lock before giving up. The default value is 50 seconds. Unit: Second.
innodb_max_dirty_pages_pct	75	75	No	[50-90]	This is an integer in the range from 0 to 100. The default value is 90 for the built-in InnoDB, 75 for InnoDB Plugin. The main thread in InnoDB tries to write pages from the buffer pool so that the percentage of dirty (not yet written) pages will not exceed this value.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
innodb_old_blocks_pct	37	37	No	[5-95]	(InnoDB Plugin only) Specifies the approximate percentage of the InnoDB buffer pool used for the old block sublist. The range of values is 5 to 95. The default value is 37 (that is, 3/8 of the pool).
innodb_old_blocks_time	1000	1000	No	[0-1024]	(InnoDB Plugin only) Specifies how long in milliseconds (ms) a block inserted into the old sublist must stay there after its first access before it can be moved to the new sublist. The default value is 0: A block inserted into the old sublist moves immediately to the new sublist the first time it is accessed, no matter how soon after insertion the access occurs. If the value is greater than 0, blocks remain in the old sublist until an access occurs at least that many ms after the first access. Unit: ms.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
innodb_open_files	3000	3000	Yes	[1-8192]	This variable is relevant only if you use multiple InnoDB tablespaces. It specifies the maximum number of .ibd files that MySQL can keep open at one time. The minimum value is 10. The default value is 300.
innodb_purge_batch_size	300	300	Yes	[1-5000]	The granularity of changes, expressed in units of redo log records, that trigger a purge operation, flushing the changed buffer pool blocks to disk.
innodb_purge_threads	1	1	Yes	[1-32]	The number of background threads devoted to the InnoDB purge operation.
innodb_read_ahead_threshold	56	56	No	[0-64]	(InnoDB Plugin only) Controls the sensitivity of linear read-ahead that InnoDB uses to prefetch pages into the buffer pool. If InnoDB reads at least innodb_read_ahead_threshold pages sequentially from an extent (64 pages), it initiates an asynchronous read for the entire following extent.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
innodb_read_io_threads	4	4	Yes	[1-64]	(InnoDB Plugin only) The number of I/O threads for read operations in InnoDB. The default value is 4.
innodb_rollback_on_timeout	OFF	OFF	Yes	[OFF|ON]	InnoDB rolls back only the last statement on a transaction timeout by default. If --innodb_rollback_on_timeout is specified, a transaction timeout causes InnoDB to abort and roll back the entire transaction (the same behavior as in MySQL 4.1). This variable was added in MySQL 5.1.15

Parameter	Default value	Running parameter value	Restart required	Value range	Description
innodb_stats_method	nulls_equal	nulls_equal	No	[nulls_equal| nulls_unequal| nulls_ignored]	How the server treats NULL values when collecting statistics about the distribution of index values for InnoDB tables. This variable has three possible values, nulls_equal, nulls_unequal, and nulls_ignored. For nulls_equal, all NULL index values are considered equal and form a single value group that has a size equal to the number of NULL values. For nulls_unequal, NULL values are considered unequal, and each NULL forms a distinct value group of size 1. For nulls_ignored, NULL values are ignored.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
innodb_stats_on_metadata	OFF	OFF	No	[ON|OFF]	When this variable is enabled (which is the default, as before the variable was created), InnoDB updates statistics during metadata statements such as SHOW TABLE STATUS or SHOW INDEX, or when accessing the INFORMATION_SCHEMA tables TABLES or STATISTICS. (These updates are similar to what happens for ANALYZE TABLE.) When disabled, InnoDB does not update statistics during these operations. Disabling this variable can improve access speed for schemas that have a large number of tables or indexes. It can also improve the stability of execution plans for queries that involve InnoDB tables.
innodb_stats_sample_pages	8	8	No	[1-4294967296]	(InnoDB Plugin only) The number of index pages to sample for index distribution statistics such as are calculated by ANALYZE TABLE. The default value is 8.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
innodb_strict_mode	OFF	OFF	No	[ON|OFF]	(InnoDB Plugin only) Whether InnoDB returns errors rather than warnings for certain conditions. This is analogous to strict SQL mode. The default value is OFF. See InnoDB Strict Mode for a list of the conditions that are affected.
innodb_table_locks	ON	ON	No	[ON|OFF]	If autocommit = 0, InnoDB honors LOCK TABLES; MySQL does not return from LOCK TABLES ... WRITE until all other threads have released all their locks to the table. The default value of innodb_table_locks is 1, which means that LOCK TABLES causes InnoDB to lock a table internally if autocommit = 0.
innodb_thread_concurrency	0	0	No	[0-128]	InnoDB tries to keep the number of operating system threads concurrently inside InnoDB less than or equal to the limit given by this variable. Once the number of threads reaches this limit, additional threads are placed into a wait state within a FIFO queue for execution.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
innodb_thread_sleep_delay	10000	10000	No	[1-3600000]	How long InnoDB threads sleep before joining the InnoDB queue, in microseconds. The default value is 10,000. A value of 0 disables sleep.Unit:ms
innodb_write_io_threads	4	4	Yes	[1-64]	(InnoDB Plugin only) The number of I/O threads for write operations in InnoDB. The default value is 4.
interactive_timeout	7200	7200	No	[10-86400]	The number of seconds the server waits for activity on an interactive connection before closing it. An interactive client is defined as a client that uses the CLIENT_INTERACTIVE option to mysql_real_connect(). Unit:second.
join_buffer_size	432 KB	432 KB	No	[128-4294967295]	The minimum size of the buffer that is used for plain index scans, range index scans, and joins that do not use indexes and thus perform full table scans.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
key_cache_age_threshold	300	300	No	[100-4294967295]	This value controls the demotion of buffers from the hot sublist of a key cache to the warm sublist. Lower values cause demotion to happen more quickly. The minimum value is 100. The default value is 300.Unit:Second.
key_cache_block_size	1024	1024	No	[512-16384]	The size in bytes of blocks in the key cache. The default value is 1024.Unit:Byte.
key_cache_division_limit	100	100	No	[1-100]	The division point between the hot and warm sublists of the key cache buffer list. The value is the percentage of the buffer list to use for the warm sublist. Permissible values range from 1 to 100. The default value is 100.
log_queries_not_using_indexes	OFF	OFF	No	[ON|OFF]	If a query takes longer than this many seconds, the server increments the Slow_queries status variable. If the slow query log is enabled, the query is logged to the slow query log file; Unit: Second.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
long_query_time	1	1	No	[0.03-10]	If a query takes longer than this many seconds, the server increments the Slow_queries status variable. If the slow query log is enabled, the query is logged to the slow query log file;Unit: Second.
loose_max_statement_time	0	0	No	[0-4294967295]	statement be interrupted if the executing time exceeds this value.
loose_rds_indexstat	OFF	OFF	No	[ON|OFF]	If ON, start to collect index information.
loose_rds_max_tmp_disk_space	10737418240	10737418240	No	[10737418240-10737418240]	RDS maximum temp disk space.
loose_rds_tablestat	OFF	OFF	No	[ON|OFF]	RDS table statistics.
loose_rds_threads_running_high_watermark	50000	50000	No	[0-50000]	Max concurrency allowed for SELECT.
loose_tokudb_buffer_pool_ratio	0	0	Yes	[0-100]	TokuDB buffer pool size ratio.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
low_priority_updates	0	0	No	[0|1]	If set to 1, all INSERT, UPDATE, DELETE, and LOCK TABLE WRITE statements wait until there is no pending SELECT or LOCK TABLE READ on the affected table. This affects only storage engines that use only table-level locking (such as MyISAM, MEMORY, and MERGE). This variable previously was named sql_low_priority_updates.
max_allowed_packet	1024 MB	1024 MB	No	[16384-1073741824]	The maximum size of one packet or any generated/intermediate string. Unit: Byte.
max_connect_errors	100	100	No	[1-4294967295]	If more than this many successive connection requests from a host are interrupted without a successful connection, the server blocks that host from further connections. You can unblock blocked hosts by flushing the host cache.
max_heap_table_size	16777216	16777216	No	[16384-4294967295]	This variable sets the maximum size to which user-created MEMORY tables are permitted to grow.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
myisam_sort_buffer_size	262144	262144	No	[262144-16777216]	The size of the buffer that is allocated when sorting MyISAM indexes during a REPAIR TABLE or when creating indexes with CREATE INDEX or ALTER TABLE.
net_read_timeout	30	30	No	[1-31536000]	The number of seconds to wait for more data from a connection before aborting the read.
net_retry_count	10	10	No	[1-4294967295]	If a read or write on a communication port is interrupted, retry this many times before giving up.
net_write_timeout	60	60	No	[1-31536000]	The number of seconds to wait for a block to be written to a connection before aborting the write.
open_files_limit	65535	65535	Yes	[4000-65535]	The number of files that the operating system permits mysqld to open. The value of this variable at runtime is the real value permitted by the system and might be different from the value you specify at server startup. The value is 0 on systems where MySQL cannot change the number of open files.
performance_schema	OFF	OFF	Yes	[ON|OFF]	Enable performance_schema or not.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
query_alloc_block_size	8192	8192	No	[1024-16384]	The allocation size of memory blocks that are allocated for objects created during statement parsing and execution. Unit: Byte.
query_cache_limit	1048576	1048576	No	[1-1048576]	Do not cache results that are larger than this number of bytes. The default value is 1MB.
query_cache_size	3145728	3145728	No	[0-104857600]	The amount of memory allocated for caching query results. The default value is 0, which disables the query cache. The permissible values are multiples of 1024; other values are rounded down to the nearest multiple. Unit: Byte.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
query_cache_type	0	0	Yes	[0|1|2]	Set the query cache type. Setting the GLOBAL value sets the type for all clients that connect thereafter. Individual clients can set the SESSION value to affect their own use of the query cache. Possible values are shown in the following table. · 0: Do not cache results in or retrieve results from the query cache. Note that this does not deallocate the query cache buffer. To do that, you should set query_cache_size to 0. · 1: Cache all cacheable query results except for those that begin with SELECT SQL_NO_CACHE. · 2: Cache results.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
query_cache_wlock_invalidate	OFF	OFF	No	[ON|OFF]	Normally, when one client acquires a WRITE lock on a MyISAM table, other clients are not blocked from issuing statements that read from the table if the query results are present in the query cache. Setting this variable to 1 causes acquisition of a WRITE lock for a table to invalidate any queries in the query cache that refer to the table. This forces other clients that attempt to access the table to wait while the lock is in effect.
query_prealloc_size	8192	8192	No	[8192-1048576]	The size of the persistent buffer used for statement parsing and execution. This buffer is not freed between statements. If you are running complex queries, a larger query_prealloc_size value might be helpful in improving performance, because it can reduce the need for the server to perform memory allocation during query execution operations. Unit: Byte.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
slow_launch_time	2	2	No	[1-1024]	If creating a thread takes longer than this many seconds, the server increments the Slow_launch_threads status variable.
sort_buffer_size	848 KB	848 KB	No	[32768-4294967295]	Each session that must perform a sort allocates a buffer of this size.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
sql_mode	\s	\s	No	(Support space and | REAL_AS_FLOAT| PIPES_AS_CONCAT| ANSI_QUOTES |IGNORE_SPACE|ONLY_FULL_GROUP_BY| NO_UNSIGNED_SUBTRACTION| NO_DIR_IN_CREATE| POSTGRESQL| ORACLE|MSSQL |DB2|MAXDB| NO_KEY_OPTIONS| NO_TABLE_OPTIONS| NO_FIELD_OPTIONS| MYSQL323| MYSQL40|ANSI |NO_AUTO_VALUE_ON_ZERO |NO_BACKSLASH_ESCAPES |STRICT_TRANS_TABLES| STRICT_ALL_TABLES| NO_ZERO_IN_DATE| NO_ZERO_DATE |ALLOW_INVALID_DATES| ERROR_FOR_DIVISION_BY_ZERO| TRADITIONAL	Modes define what SQL syntax MySQL should support and what kind of data validation checks it should perform.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
table_definition_cache	512	512	No	[400-80480]	The number of table definitions (from .frm files) that can be stored in the definition cache. If you use a large number of tables, you can create a large table definition cache to speed up opening of tables. The table definition cache takes less space and does not use file descriptors, unlike the normal table cache. The minimum and default values are both 400.
table_open_cache	2000	2000	No	[1-524288]	The stack size of each thread.
thread_stack	262144	262144	Yes	[131072-1844674407 3709551615]	The stack size of each thread.
tmp_table_size	2097152	2097152	No	[262144-67108864]	The maximum size of internal in-memory temporary tables.
wait_timeout	86400	86400	No	[60-259200]	The number of seconds the server waits for activity on a noninteractive connection before closing it.

Table 9-19: Modifiable MySQL 5.7 instance parameters

Parameter	Default value	Running parameter value	Restart required	Value range	Description
auto_increment_increment	1	1	No	[1-65535]	auto_increment_increment and auto_increment_offset are intended for use with master-to-master replication, and can be used to control the operation of AUTO_INCREMENT columns. Both variables have global and session values, and each can assume an integer value between 1 and 65,535 inclusive. Setting the value of either of these two variables to 0 causes its value to be set to 1 instead. Attempting to set the value of either of these two variables to an integer greater than 65,535 or less than 0 causes its value to be set to 65,535 instead. Attempting to set the value of auto_increment_increment or auto_increment_offset to a noninteger value produces an error, and the actual value of the variable remains unchanged.
auto_increment_offset	1	1	No	[1-65535]	determines the starting point for the AUTO_INCREMENT column value.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
back_log	3000	3000	Yes	[0-65535]	The number of outstanding connection requests that MySQL can have.
binlog_cache_size	128 KB	128 KB	No	[4096-16777216]	The size of the cache to hold changes to the binary log during a transaction.
binlog_checksum	CRC32	CRC32	Yes	[CRC32|NONE]	The master to write a checksum for each event in the binary log.
binlog_order_commits	ON	ON	No	[ON|OFF]	When this variable is enabled on a master (the default), transactions are externalized in the same order as they are written to the binary log.
binlog_row_image	full	full	No	[full|minimal]	Binlog save every column or actually required column in binlog images.
binlog_stmt_cache_size	32768	32768	No	[4096-16777216]	The size of the statement cache for updates to non-transactional engines for the binary log.
block_encryption_mode	"aes-128-ecb"	"aes-128-ecb"	No	["aes-128-ecb"|"aes-192-ecb"|"aes-256-ecb"|"aes-128-cbc"|"aes-192-cbc"|"aes-256-cbc"]	This variable controls the block encryption mode for block-based algorithms such as AES. It affects encryption for AES_ENCRYPT() and AES_DECRYPT().
character_set_server	utf8	utf8	Yes	[utf8|latin1|gbk|utf8mb4]	The server's default character set.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
concurrent_insert	1	1	No	[0|1|2]	NEVER-Disables concurrent inserts; AUTO-(Default) Enables concurrent insert for MyISAM tables that do not have holes; ALWAYS -Enables concurrent inserts for all MyISAM tables, even those that have holes. For a table with a hole, new rows are inserted at the end of the table if it is in use by another thread . Otherwise, MySQL acquires a normal write lock and inserts the row into the hole.
connect_timeout	10	10	No	[1-3600]	The number of seconds that the mysqld server waits for a connect packet before responding with Bad handshake. The default value is 10 seconds as of MySQL 5.1.23 and 5 seconds before that.Increasing the connect_timeout value might help if clients frequently encounter errors of the form Lost connection to MySQL server at 'XXX', system error: errno.
default_storage_engine	InnoDB	TokuDB	Yes	[InnoDB| TokuDB|innodb |tokudb]	The default storage engine for new tables.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
default_time_zone	SYSTEM	SYSTEM	Yes	[SYSTEM|-12:00|-11:00|-10:00|-9:00|-8:00|-7:00|-6:00|-5:00|-4:00|-3:00|-2:00|-1:00|+0:00|+1:00|+2:00|+3:00|+4:00|+5:00|+5:30|+6:00|+6:30|+7:00|+8:00|+9:00|+10:00|+11:00|+12:00|+13:00]	The default time zone for the database.
default_week_format	0	0	No	[0-7]	The default mode value to use for the WEEK() function.
delayed_insert_limit	100	100	No	[1-4294967295]	After inserting delayed_insert_limit delayed rows, the INSERT DELAYED handler thread checks whether there are any SELECT statements pending. If so, it permits them to execute before continuing to insert delayed rows.
delayed_insert_timeout	300	300	No	[1-3600]	How many seconds an INSERT DELAYED handler thread should wait for INSERT statements before terminating.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
delayed_queue_size	1000	1000	No	[1-4294967295]	This is a per-table limit on the number of rows to queue when handling INSERT DELAYED statements. If the queue becomes full, any client that issues an INSERT DELAYED statement waits until there is room in the queue again.
delay_key_write	ON	ON	No	[ON|OFF|ALL]	This option applies only to MyISAM tables. It can have one of the following values to affect handling of the DELAY_KEY_WRITE table option that can be used in CREATE TABLE statements.
div_precision_increment	4	4	No	[0-30]	This variable indicates the number of digits by which to increase the scale of the result of division operations performed with the / operator. The default value is 4. The minimum and maximum values are 0 and 30, respectively. The following example illustrates the effect of increasing the default value.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
eq_range_index_dive_limit	10	10	No	[1-200]	The optimizer will use existing index statistics instead of doing index dives for equality ranges if the number of equality ranges for the index is larger than or equal to this number. If set to 0, index dives are always used.
event_scheduler	OFF	OFF	No	[ON|OFF]	Enable the event scheduler. Possible values are ON, OFF, and DISABLED (keep the event scheduler completely deactivated, it cannot be activated run-time).
ft_min_word_len	4	4	Yes	[1-3600]	The minimum length of the word to be included in a FULLTEXT index.
ft_query_expansion_limit	20	20	Yes	[0-1000]	The number of top matches to use for full-text searches performed using WITH QUERY EXPANSION.
group_concat_max_len	1024	1024	No	[4-18446744073709547520]	The maximum permitted result length in bytes for the GROUP_CONCAT() function. The default is 1024, Unit:Byte.
innodb_adaptive_hash_index	ON	ON	No	[ON|OFF]	Enable InnoDB adaptive hash index (enabled by default). Disable with --skip-innodb-adaptive-hash-index.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
innodb_autoinc_lock_mode	1	1	Yes	[0|1|2]	The number of threads that can enter InnoDB concurrently is determined by the innodb_thread_concurrency variable.
innodb_concurrency_tickets	5000	5000	No	[1-4294967295]	The number of threads that can enter InnoDB concurrently is determined by the innodb_thread_concurrency variable.
innodb_ft_max_token_size	84	84	Yes	[10-84]	Maximum length of words that are stored in an InnoDB FULLTEXT index.
innodb_ft_min_token_size	3	3	Yes	[0-16]	Minimum length of words that are stored in an InnoDB FULLTEXT index.
innodb_large_prefix	OFF	OFF	No	[ON|OFF]	Enable this option to allow index key prefixes longer than 767 bytes (up to 3072 bytes), for InnoDB tables that use the DYNAMIC and COMPRESSED row formats.
innodb_lock_wait_timeout	50	50	No	[1-1073741824]	The timeout in seconds an InnoDB transaction may wait for a row lock before giving up. The default value is 50 seconds. Unit: Second.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
innodb_max_dirty_pages_pct	75	75	No	[50-90]	This is an integer in the range from 0 to 100. The default value is 90 for the built-in InnoDB, 75 for InnoDB Plugin. The main thread in InnoDB tries to write pages from the buffer pool so that the percentage of dirty (not yet written) pages will not exceed this value.
innodb_old_blocks_pct	37	37	No	[5-95]	(InnoDB Plugin only) Specifies the approximate percentage of the InnoDB buffer pool used for the old block sublist. The range of values is 5 to 95. The default value is 37 (that is, 3/8 of the pool).

Parameter	Default value	Running parameter value	Restart required	Value range	Description
innodb_old_blocks_time	1000	1000	No	[0-1024]	(InnoDB Plugin only) Specifies how long in milliseconds (ms) a block inserted into the old sublist must stay there after its first access before it can be moved to the new sublist. The default value is 0: A block inserted into the old sublist moves immediately to the new sublist the first time it is accessed, no matter how soon after insertion the access occurs. If the value is greater than 0, blocks remain in the old sublist until an access occurs at least that many ms after the first access. Unit: ms.
innodb_online_alter_log_max_size	134217728	134217728	No	[134217728-2147483647]	Maximum modification log file size for online index creation.
innodb_open_files	3000	3000	Yes	[1-8192]	This variable is relevant only if you use multiple InnoDB tablespaces. It specifies the maximum number of .ibd files that MySQL can keep open at one time. The minimum value is 10. The default value is 300.
innodb_print_all_deadlocks	OFF	OFF	No	[OFF|ON]	Print all deadlocks to MySQL error log (off by default).

Parameter	Default value	Running parameter value	Restart required	Value range	Description
innodb_purge_batch_size	300	300	Yes	[1-5000]	The granularity of changes, expressed in units of redo log records, that trigger a purge operation, flushing the changed buffer pool blocks to disk.
innodb_purge_threads	1	1	Yes	[1-32]	The number of background threads devoted to the InnoDB purge operation.
innodb_read_ahead_threshold	56	56	No	[0-64]	(InnoDB Plugin only) Controls the sensitivity of linear read-ahead that InnoDB uses to prefetch pages into the buffer pool. If InnoDB reads at least innodb_read_ahead_threshold pages sequentially from an extent (64 pages), it initiates an asynchronous read for the entire following extent.
innodb_read_io_threads	4	4	Yes	[1-64]	(InnoDB Plugin only) The number of I/O threads for read operations in InnoDB. The default value is 4.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
innodb_rollback_on_timeout	OFF	OFF	Yes	[OFF|ON]	InnoDB rolls back only the last statement on a transaction timeout by default. If --innodb_rollback_on_timeout is specified, a transaction timeout causes InnoDB to abort and roll back the entire transaction (the same behavior as in MySQL 4.1). This variable was added in MySQL 5.1.15
innodb_stats_method	nulls_equal	nulls_equal	No	[nulls_equal|nulls_unequal|nulls_ignored]	How the server treats NULL values when collecting statistics about the distribution of index values for InnoDB tables. This variable has three possible values, nulls_equal, nulls_unequal, and nulls_ignored. For nulls_equal, all NULL index values are considered equal and form a single value group that has a size equal to the number of NULL values. For nulls_unequal, NULL values are considered unequal, and each NULL forms a distinct value group of size 1. For nulls_ignored, NULL values are ignored.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
innodb_stats_on_metadata	OFF	OFF	No	[ON|OFF]	When this variable is enabled (which is the default, as before the variable was created), InnoDB updates statistics during metadata statements such as SHOW TABLE STATUS or SHOW INDEX, or when accessing the INFORMATION_SCHEMA tables TABLES or STATISTICS. (These updates are similar to what happens for ANALYZE TABLE.) When disabled, InnoDB does not update statistics during these operations. Disabling this variable can improve access speed for schemas that have a large number of tables or indexes. It can also improve the stability of execution plans for queries that involve InnoDB tables.
innodb_stats_sample_pages	8	8	No	[1-4294967296]	(InnoDB Plugin only) The number of index pages to sample for index distribution statistics such as are calculated by ANALYZE TABLE. The default value is 8.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
innodb_strict_mode	OFF	OFF	No	[ON|OFF]	(InnoDB Plugin only) Whether InnoDB returns errors rather than warnings for certain conditions. This is analogous to strict SQL mode. The default value is OFF. See InnoDB Strict Mode for a list of the conditions that are affected.
innodb_table_locks	ON	ON	No	[ON|OFF]	If autocommit = 0, InnoDB honors LOCK TABLES; MySQL does not return from LOCK TABLES ... WRITE until all other threads have released all their locks to the table. The default value of innodb_table_locks is 1, which means that LOCK TABLES causes InnoDB to lock a table internally if autocommit = 0.
innodb_thread_concurrency	0	0	No	[0-128]	InnoDB tries to keep the number of operating system threads concurrently inside InnoDB less than or equal to the limit given by this variable. Once the number of threads reaches this limit, additional threads are placed into a wait state within a FIFO queue for execution.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
innodb_thread_sleep_delay	10000	10000	No	[1-3600000]	How long InnoDB threads sleep before joining the InnoDB queue, in microseconds. The default value is 10,000. A value of 0 disables sleep.Unit:ms
innodb_write_io_threads	4	4	Yes	[1-64]	(InnoDB Plugin only) The number of I/O threads for write operations in InnoDB. The default value is 4.
interactive_timeout	7200	7200	No	[10-86400]	The number of seconds the server waits for activity on an interactive connection before closing it. An interactive client is defined as a client that uses the CLIENT_INTERACTIVE option to mysql_real_connect(). Unit:second.
join_buffer_size	432 KB	432 KB	No	[128-4294967295]	The minimum size of the buffer that is used for plain index scans, range index scans, and joins that do not use indexes and thus perform full table scans.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
key_cache_age_threshold	300	300	No	[100-4294967295]	This value controls the demotion of buffers from the hot sublist of a key cache to the warm sublist. Lower values cause demotion to happen more quickly. The minimum value is 100. The default value is 300.Unit:Second.
key_cache_block_size	1024	1024	No	[512-16384]	The size in bytes of blocks in the key cache. The default value is 1024.Unit:Byte.
key_cache_division_limit	100	100	No	[1-100]	The division point between the hot and warm sublists of the key cache buffer list. The value is the percentage of the buffer list to use for the warm sublist. Permissible values range from 1 to 100. The default value is 100.
log_queries_not_using_indexes	OFF	OFF	No	[ON|OFF]	If a query takes longer than this many seconds, the server increments the Slow_queries status variable. If the slow query log is enabled, the query is logged to the slow query log file; Unit: Second.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
long_query_time	1	1	No	[0.03-10]	If a query takes longer than this many seconds, the server increments the Slow_queries status variable. If the slow query log is enabled, the query is logged to the slow query log file;Unit: Second.
loose_opt_rds_enable_show_slave_lag	ON	ON	No	bool	if ON, the 'show slave lag' command is allowed, default is (OFF).
loose_opt_rds_last_error_gtid	ON	ON	No	bool	if ON, show Last_SQL_Error_Gtid in show slave status.
loose_rds_check_core_file_enabled	ON	ON	No	bool	enable check core file.
loose_rds_kill_connections	20	20	No	[0, 18446744073709551615]	The extra connection for user in rds_user_with_kill_option.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
low_priority_updates	0	0	No	[0|1]	If set to 1, all INSERT, UPDATE, DELETE, and LOCK TABLE WRITE statements wait until there is no pending SELECT or LOCK TABLE READ on the affected table. This affects only storage engines that use only table-level locking (such as MyISAM, MEMORY, and MERGE). This variable previously was named sql_low_priority_updates.
max_connect_errors	100	100	No	[1-4294967295]	If more than this many successive connection requests from a host are interrupted without a successful connection, the server blocks that host from further connections. You can unblock blocked hosts by flushing the host cache.
max_heap_table_size	16777216	16777216	No	[16384-4294967295]	This variable sets the maximum size to which user-created MEMORY tables are permitted to grow.
max_length_for_sort_data	1024	1024	No	[0-838860]	Max number of bytes in sorted records.
max_prepared_stmt_count	16382	16382	No	[0-1048576]	Maximum number of prepared statements in the server.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
max_write_lock_count	102400	102400	No	[1-102400]	After this many write locks, allow some read locks to run in between.
myisam_sort_buffer_size	262144	262144	No	[262144-16777216]	The size of the buffer that is allocated when sorting MyISAM indexes during a REPAIR TABLE or when creating indexes with CREATE INDEX or ALTER TABLE.
net_read_timeout	30	30	No	[1-31536000]	The number of seconds to wait for more data from a connection before aborting the read.
net_retry_count	10	10	No	[1-4294967295]	If a read or write on a communication port is interrupted, retry this many times before giving up.
net_write_timeout	60	60	No	[1-31536000]	The number of seconds to wait for a block to be written to a connection before aborting the write.
ngram_token_size	2	2	Yes	[0-20]	Defines the n-gram token size for the n-gram full-text parser. The ngram_token_size option is read-only and can only be modified at startup. The default value is 2 (bigram). The maximum value is 10.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
open_files_limit	65535	65535	Yes	[4000-65535]	The number of files that the operating system permits mysqld to open. The value of this variable at runtime is the real value permitted by the system and might be different from the value you specify at server startup. The value is 0 on systems where MySQL cannot change the number of open files.
performance_schema	OFF	OFF	Yes	[ON|OFF]	Enable performance_schema or not.
query_alloc_block_size	8192	8192	No	[1024-16384]	The allocation size of memory blocks that are allocated for objects created during statement parsing and execution. Unit: Byte.
query_cache_limit	1048576	1048576	No	[1-1048576]	Do not cache results that are larger than this number of bytes. The default value is 1MB.
query_cache_size	3145728	3145728	No	[0-104857600]	The amount of memory allocated for caching query results. The default value is 0, which disables the query cache. The permissible values are multiples of 1024; other values are rounded down to the nearest multiple. Unit: Byte.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
query_cache_type	0	0	Yes	[0|1|2]	Set the query cache type. Setting the GLOBAL value sets the type for all clients that connect thereafter. Individual clients can set the SESSION value to affect their own use of the query cache. Possible values are shown in the following table. · 0: Do not cache results in or retrieve results from the query cache. Note that this does not deallocate the query cache buffer. To do that, you should set query_cache_size to 0. · 1: Cache all cacheable query results except for those that begin with SELECT SQL_NO_CACHE. · 2: Cache results.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
query_prealloc_size	8192	8192	No	[8192-1048576]	The size of the persistent buffer used for statement parsing and execution. This buffer is not freed between statements. If you are running complex queries, a larger query_prealloc_size value might be helpful in improving performance, because it can reduce the need for the server to perform memory allocation during query execution operations. Unit: Byte.
rds_reserved_connections	512	512	No	[0-512]	The reserved connection for maintain user.
slave_parallel_type	LOGICAL_CLOCK	LOGICAL_CLOCK	No	DATABASE, LOGICAL_CLOCK	Specifies if the slave will use database partitioning or information from master to parallelize transactions.(Default: DATABASE).
slow_launch_time	2	2	No	[1-1024]	If creating a thread takes longer than this many seconds, the server increments the Slow_launch_threads status variable.
sort_buffer_size	848 KB	848 KB	No	[32768-4294967295]	Each session that must perform a sort allocates a buffer of this size.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
sql_mode	\s	\s	No	(Support space and | REAL_AS_FLOAT| PIPES_AS_CONCAT| ANSI_QUOTES |IGNORE_SPACE|ONLY_FULL_GROUP_BY| NO_UNSIGNED_SUBTRACTION| NO_DIR_IN_CREATE| POSTGRESQL| ORACLE|MSSQL |DB2|MAXDB| NO_KEY_OPTIONS| NO_TABLE_OPTIONS| NO_FIELD_OPTIONS| MYSQL323| MYSQL40|ANSI |NO_AUTO_VALUE_ON_ZERO |NO_BACKSLASH_ESCAPES |STRICT_TRANS_TABLES| STRICT_ALL_TABLES| NO_ZERO_IN_DATE| NO_ZERO_DATE |ALLOW_INVALID_DATES| ERROR_FOR_DIVISION_BY_ZERO| TRADITIONAL	Modes define what SQL syntax MySQL should support and what kind of data validation checks it should perform.

Parameter	Default value	Running parameter value	Restart required	Value range	Description
table_definition_cache	512	512	No	[400-80480]	The number of table definitions (from .frm files) that can be stored in the definition cache. If you use a large number of tables, you can create a large table definition cache to speed up opening of tables. The table definition cache takes less space and does not use file descriptors, unlike the normal table cache. The minimum and default values are both 400.
table_open_cache	2000	2000	No	[1-524288]	The stack size of each thread.
tmp_table_size	2097152	2097152	No	[262144-67108864]	The maximum size of internal in-memory temporary tables.
transaction_isolation	READ-COMMITTED	READ-COMMITTED	No	[READ-UNCOMMITTED|READ-COMMITTED|REPEATABLE-READ|SERIALIZABLE]	Default transaction isolation level.
wait_timeout	86400	86400	No	[60-259200]	The number of seconds the server waits for activity on a noninteractive connection before closing it.

9.4.10.2 Best practices for MySQL instance parameter optimization

9.4.10.2.1 Overview

You can optimize MySQL parameters for RDS instances. This topic describes the best practices for modifiable and unmodifiable MySQL parameters. It also describes how to optimize modifiable parameters to improve instance performance.

9.4.10.2.2 Unmodifiable MySQL instance parameters

This topic describes the RDS for MySQL parameters that cannot be modified.

Different types of RDS for MySQL instances have different specifications for the maximum number of connections and memory size. Therefore, parameters related to the instance type, such as connections and memory, cannot be modified. You can resolve connection or memory bottlenecks by using the following methods:

- **Memory bottleneck: An out of memory (OOM) error occurs in an instance. This can be resolved with a primary/secondary failover.**
- **Connection bottleneck: An application cannot establish connections to the database. You can upgrade the instance type, optimize the application, or slow SQL statements.**

To ensure the security of primary and secondary instance data, the following security-related parameters cannot be modified: innodb_flush_log_at_trx_commit, sync_binlog, gtid_mode, semi_sync, and binlog_format.

9.4.10.2.3 Modifiable MySQL instance parameters

This topic describes the RDS for MySQL parameters that can be modified.

Unmodifiable MySQL instance parameters **lists the MySQL instance parameters that cannot be modified. The parameters of RDS for MySQL have been optimized by DBA and source code teams. This helps you run your database without the need to adjust any parameters. For more information about the modifiable RDS for MySQL instance parameters, see** *Modifiable MySQL instance parameters***. These parameters are applicable in most scenarios. They only need to be adjusted in certain cases. Example:**

- **If you use TokuDB, you must adjust the percentage of the memory available for the engine by using the tokudb_buffer_pool_ratio parameter.**

- **If your applications require a relatively long lock timeout period, you must adjust the innodb_lock_wait_timeout parameter.**

9.4.10.2.4 How to configure parameters

This topic describes how to configure important parameters in the RDS for MySQL console. If these parameters are incorrectly configured, your instances may encounter performance problems or applications may report errors.

open_files_limit

Function: This parameter controls the number of file handles that can be simultaneously enabled by each MySQL instance. When a database table is opened, it consumes a file handle allocated to the instance. RDS for MySQL sets open_files_limit to 8,192 when initializing an instance. When the number of consumed file handles exceeds this value, errors are returned for all database requests.

Note:

Access to MyISAM tables consumes file descriptors. InnoDB uses table_open_cache to manage opened tables.

Symptom: If the value of this parameter is set too low, applications may report the following error:

```
[ERROR] /mysqld: Can't open file: './mysql/user.frm' (errno: 24 -Too
many open files);
```

Suggestion: Increase the value of open_files_limit. RDS for MySQL allows you to set the maximum value to 65,535 for this parameter. We also recommend that you replace MyISAM with InnoDB.

back_log

Function: RDS for MySQL creates a thread for every connection request that it processes. If front-end applications initiate too many transient database connection requests when a thread is created, RDS for MySQL uses back_log to restrict the number of queued connection requests. If the number of connection requests in the queue exceeds the value of back_log, RDS for MySQL denies new connection requests. If you want MySQL to process a large number of transient connection requests, increase the value of back_log.

Symptom: If the value of this parameter is set too low, applications may report the following error:

```
SQLSTATE[HY000] [2002] Connection timed out;
```

Suggestion: Increase the value of back_log. The default value of this parameter has been increased from 50 to 3,000.

Notice:

You must restart your instances after you modify the parameter value.

innodb_autoinc_lock_mode

Function: In RDS for MySQL 5.1.22 and later versions, innodb_autoinc_lock_mode is introduced to InnoDB to control auto-increment locks. This parameter can be set to 0, 1, or 2. The default value is 1 in RDS for MySQL. This value indicates that InnoDB uses the lightweight mutex lock to obtain auto-increment locks in place of table-level locks. However, the load data statements (including INSERT ... SELECT and REPLACE ... SELECT statements) use auto-increment table locks. A deadlock may occur when multiple applications use this statement to load data at the same time.

Symptom: A deadlock occurs when multiple applications use the load data statements such as INSERT ... SELECT and REPLACE ... SELECT statements to load data at the same time. The following error is reported:

```
RECORD LOCKS space id xx page no xx n bits xx index PRIMARY of table
 xx.xx trx id xxx lock_mode X insert intention waiting. TABLE LOCK
table xxx.xxx trx id xxxx lock mode AUTO-INC waiting;
```

Suggestion: We recommend that you change the value of innodb_autoinc_lock_mode to 2 to enable the use of the lightweight mutex lock (only in the row mode) for all INSERT statements. This avoids auto_inc deadlocks and greatly improves the performance of the INSERT ... SELECT statement.

Note:

If you set the parameter value to 2, you must set the format of binlog to row.

query_cache_size

Function: This parameter controls the memory size of the MySQL query cache. If the query cache is enabled, MySQL locks the query cache when it executes a query.

Then, MySQL determines whether the query cache contains the queried data. If yes, MySQL directly returns results. Otherwise, MySQL proceeds to perform other operations such as engine query. The INSERT, UPDATE, and DELETE statements can invalidate the query cache and any changes to schemas and indexes. The cost of maintaining an invalid query cache is relatively high, which puts a lot of pressure on MySQL. The query cache helps improve instance performance when the database is not frequently updated. However, when data is frequently written to several tables in the database, the query cache lock can result in frequent lock conflicts. This is because read and write operations on a specific table must wait for the query cache lock to be unlocked. This reduces the efficiency of the SELECT statement.

Symptom: The database goes through the following status: checking query cache for query, waiting for query cache lock, and storing results in query cache.

Suggestion: ApsaraDB for RDS disables the query cache by default. If the query cache is enabled and you encounter the preceding problems, you can disable the query cache. However, you can enable the query cache to solve database performance issues in some cases.

net_write_timeout

Function: This parameter sets the timeout period that ApsaraDB for RDS waits before it sends a block to a client.

Symptom: If the value of the parameter is set too low, the client may report the following error:

```
the last packet successfully received from the server was milliseconds
 ago, the last packet sent successfully to the server was milliseconds
 ago.
```

Suggestion: The default value is 60 seconds. A small value of net_write_timeout may result in frequent disconnections in cases where the network conditions are poor or it takes a long time for the client to process each block. In these cases, we recommend that you increase the value of this parameter.

tmp_table_size

Function: This parameter determines the maximum size of the internal temporary memory table. It is assigned to each thread. The actual value is the smaller one between tmp_table_size and max_heap_table_size. If the size of a temporary

memory table exceeds the value of this parameter, MySQL automatically converts the table to a disk-based MyISAM table. Avoid using temporary tables when you optimize query statements. If you need to use a temporary table, make sure that the temporary table is stored in the memory.

Symptom: If you use a temporary table for complicated SQL statements that contain GROUP BY or DISTINCT clauses, which cannot be optimized through indexes, SQL execution takes a longer time.

Suggestion: If an application involves many GROUP BY or DISTINCT clauses and the database has enough memory, you can increase the values of tmp_table_size and max_heap_table_size to improve query performance.

9.4.10.2.5 New MySQL parameters

This topic describes the new parameters of RDS for MySQL.

oose_rds_max_tmp_disk_space

Function: This parameter controls the temporary file size available for MySQL.
Default value: 10 GB.

Symptom: If the temporary file size exceeds the limit indicated by this parameter, applications may report the following error:

```
The table '/home/mysql/dataxxx/tmp/#sql_2db3_1' is full.
```

Suggestion: Evaluate whether you can optimize the SQL statements that cause additional temporary files using indexing or other means. If your instance has enough space, you can increase the value of this parameter to guarantee that SQL statements execute normally.

Notice:
You must restart your instances after you change the value of this parameter.

loose_tokudb_buffer_pool_ratio

Function: This parameter controls the buffer size available for TokuDB.
For example, if you set innodb_buffer_pool_size to 1,000 MB and tokudb_buffer_pool_ratio to 50 to indicate 50% of the buffer size, TokuDB tables can use up to 500 MB of the buffer space.

Suggestion: The default value is 0. If you use TokuDB in your RDS for MySQL instance, we recommend that you increase the value of this parameter to improve the access performance of TokuDB tables.

Notice:
You must restart your instances after you change the value of this parameter.

loose_max_statement_time

Function: This parameter sets a limit on how long a query can take in MySQL before timing out.

Symptom: There is no maximum query time limit by default. If this parameter is set and the query time exceeds the specified limit, the query fails with the following error:

```
ERROR 3006 (HY000): Query execution was interrupted, max_statem
ent_time exceeded
```

Suggestion: Modify this parameter if you want to control the SQL execution time (in milliseconds) for your database.

loose_rds_threads_running_high_watermark

Function: This parameter controls the maximum number of concurrent MySQL queries. For example, if you set rds_threads_running_high_watermark to 100, 100 MySQL queries can be initiated concurrently on the instance. Additional queries are denied. This parameter is used with rds_threads_running_ctl_mode (default value: select).

Suggestion: This parameter is typically used to handle burst requests and requests during peak hours to protect databases.

9.5 Account

9.5.1 Create an account

This topic describes the functions and features of accounts in classic and premier modes, and how to create accounts in different modes.

You must create an account in an RDS instance before you can use the database. ApsaraDB for RDS supports two account management modes: classic and premier . The classic mode is a management mode retained from earlier versions of

ApsaraDB for RDS. In the classic mode, databases and accounts cannot be managed through SQL. The premier mode is a management mode introduced in later versions that enables more permissions. In the premier mode, databases and accounts can be managed through SQL. We recommend that you use the premier mode for personalized and fine-grained control over database permissions.

Account modes

In the classic mode, all accounts are created by using the RDS console or API operations, instead of through SQL. All accounts are equal. There is not one account with more management permissions over others. You can use the RDS console to create and manage all accounts and databases.

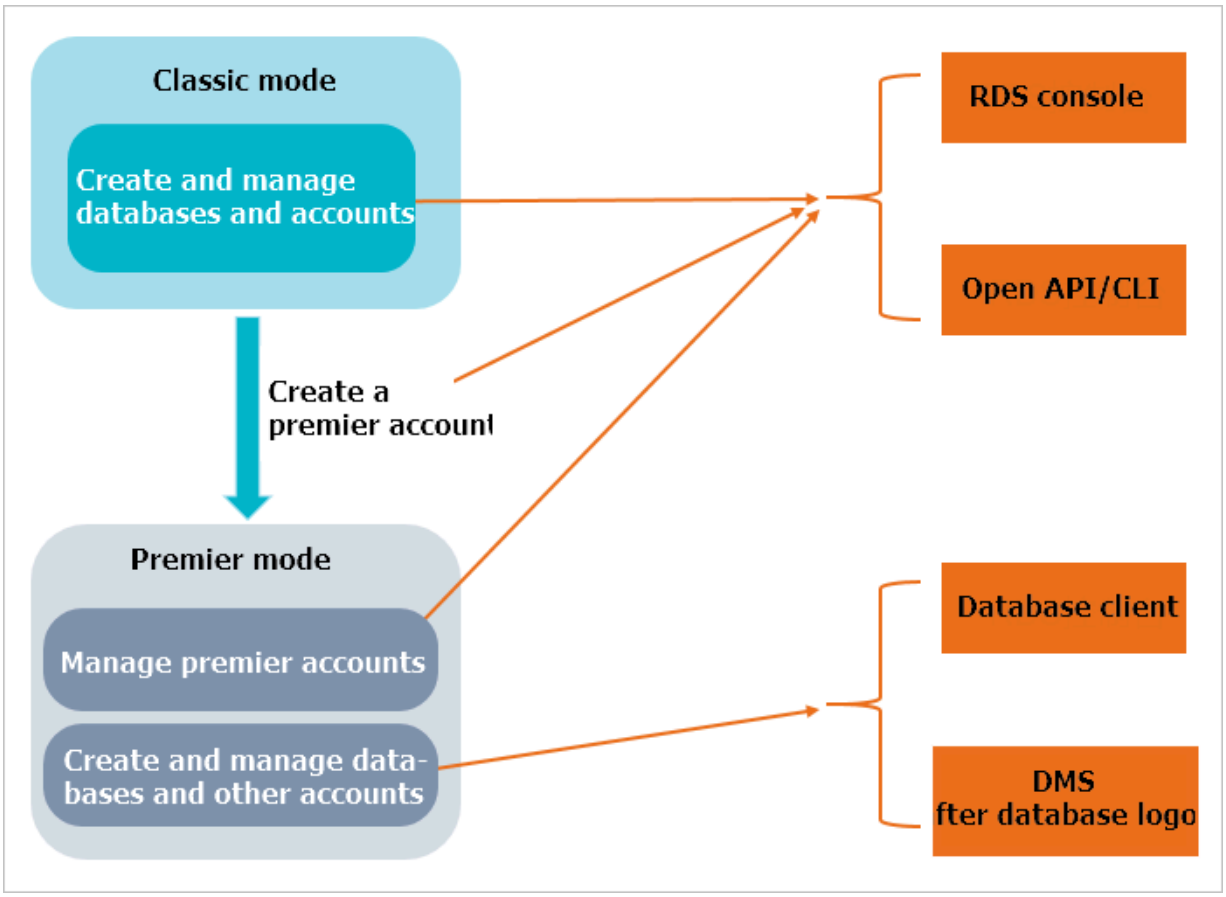
To activate the premier mode, you must create a Super Administrator premier account. You can use the RDS console or API operations to create and manage premier accounts. Log on to a database with your premier account. You can then use SQL statements or DMS to create and manage standard accounts. Run the following commands to log on to the database by using the premier account named root and create a standard account named jeffrey:

```
mysql -hxxxxxxxxx.mysql.rds.aliyuncs.com -uroot -pxxxxxx -e "
    CREATE USER 'jeffrey'@'%' IDENTIFIED BY 'password';
    CREATE DATABASE DB001;
    "
```

In the premier mode, you cannot manage databases by using the RDS console or API operations. You must use SQL statements or DMS to create and manage databases.

Figure 9-20: Difference between standard and premier accounts **shows how to create and manage databases and accounts in classic and premier modes.**

Figure 9-20: Difference between standard and premier accounts



How to create an account

Note:

- **Use service roles to create accounts and follow the principle of least privilege to assign appropriate read-only and read/write permissions to accounts. When necessary, you can create many database accounts and allow each of them to access data only relevant to their own business tasks. If an account does not need to write data to a database, assign read-only permissions to the account.**
- **Use strong passwords for database accounts and change the passwords on a regular basis.**

Procedure

- **For more information about how to create a standard account for ApsaraDB RDS for MySQL, see** *Create a standard account*.

- For more information about how to create a premier account for ApsaraDB RDS for MySQL, see *Create a privileged account*.
- For more information about how to create an account for ApsaraDB RDS for PostgreSQL, see *Create accounts and databases*.
- For more information about how to create an account for ApsaraDB RDS for PPAS, see *Create accounts and databases*.

9.5.2 Reset your password

You can use the RDS console to reset the password of your database account if you forget the password.

Context

Notice:
To ensure data security, we recommend that you change your password on a regular basis.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane of the Basic Information page, choose Database Management > Accounts.
4. On the Accounts page, click the icon in the Actions column corresponding to the account that you want to modify and choose Reset Password from the shortcut menu. In the Reset Password dialog box that appears, configure the parameters as prompted.

   Table 9-20: Password resetting parameters describes the parameter configurations.

   Table 9-20: Password resetting parameters

Parameter	Description
New Password	The password must be 6 to 32 characters in length and can contain letters, digits, and underscores (_).
Retype Password	The password must be 6 to 32 characters in length and can contain letters, digits, and underscores (_).

5. Click OK.

9.5.3 Modify account permissions

You can modify the account permissions of your instances at any time when using ApsaraDB for RDS.

Prerequisites

Permissions of standard accounts can be deleted in the console for instances without a premier account. Standard accounts can only be created and managed using SQL commands or DMS for instances with a premier account.

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane of the Basic Information page, choose Database Management > Accounts to go to the Accounts page.**
4. **On the Accounts page, click the icon in the Actions column corresponding to the account that you want to modify and choose Modify Permissions from the shortcut menu. On the page that appears, modify the account permissions as prompted.**

   Table 9-21: Account permission modification parameters **describes the parameter configurations.**

   Table 9-21: Account permission modification parameters

Parameter	Description
Available Databases	The list of databases that have been created.
Selected Databases	The databases whose permissions are to be granted to the account. You can select one of the following permissions: · Read-Only: grants the database read-only permissions to the account. · Read and Write: grants the database read/write permissions to the account.

5. **Click OK.**

9.5.4 Delete an account

You can use the console to delete standard accounts that are no longer used.

Prerequisites

Accounts can be deleted in the console for instances without a premier account. Accounts can only be deleted using SQL commands or DMS for instances with a premier account.

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane of the Basic Information page, choose Database Management > Accounts to go to the Accounts page.**
4. **On the Accounts page, click the icon in the Actions column corresponding to the account that you want to delete and choose Delete from the shortcut menu. In the Delete User message that appears, click OK.**

9.5.5 Modify the account description

You can add a description when you create an account to ease management operations. After the account is created, you can modify the description at any time.

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane of the Basic Information page, choose Database Management > Accounts.**
4. **On the Accounts page, click the icon in the Actions column corresponding to the account that you want to modify and choose Change Description from the shortcut menu. In the Change Account Description dialog box that appears, set Description.**
5. **Click OK.**

9.6 Database

9.6.1 Create a database

After you create an RDS instance and configure the whitelist, you need to create a database and an account in the instance.

Prerequisites

Databases can be created in the console for instances without a premier account. Databases can only be created using SQL commands or DMS for instances with a premier account.

Context

To migrate the local database to ApsaraDB for RDS, you must create a database and an account in the RDS instance identical to those in the local database. Use service roles to create accounts and follow the principle of least privilege to assign appropriate read-only and read/write permissions to accounts. When necessary, you can create many database accounts and allow each of them to access data only relevant to their own business tasks.

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane of the Basic Information page, choose Database Management > Databases.**
4. **On the Databases page, click Create Database. On the Create Database page that appears, configure the parameters as prompted.**

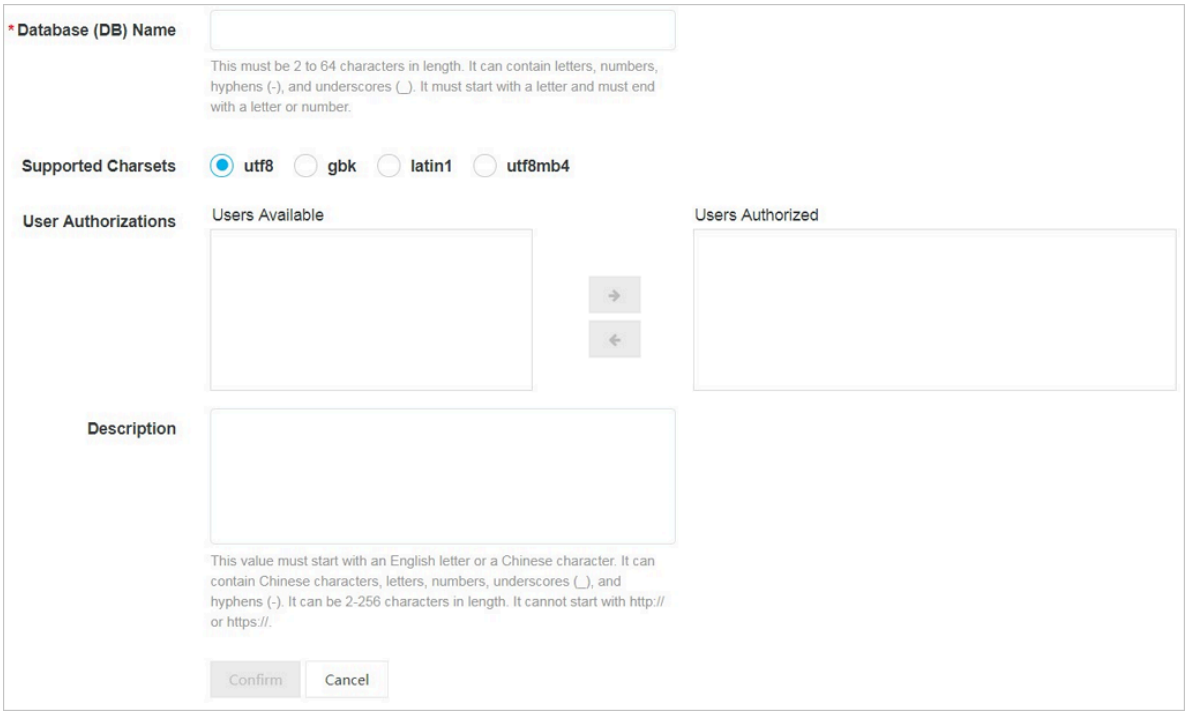
   Table 9-22: Database creation parameters **describes the parameter configurations.**

   Table 9-22: Database creation parameters

Parameter	Description
Database (DB) Name	The database name. The naming conventions are as follows: · It must start with a lowercase letter and end with a lowercase letter or digit. · It can contain lowercase letters, digits, underscores (_), and hyphens (-). · It must be 2 to 64 characters in length. · It cannot contain reserved keywords. For more information about the reserved keywords, see the relevant document in the official website of the corresponding engine.

Parameter	Description
Supported Character Set	The character sets that are supported by the database, including: · utf8 · gbk · latin1 · utf8mb4
User Authorizations	· Select an account to grant read-only or read/write permissions. · The permissions on the database can be granted only to standard accounts. · By default, the premier account is authorized to use the database.
Description	The description of the database. The rules are as follows: · It must start with a lowercase letter. · It can contain lowercase letters, digits, underscores (_), and hyphens (-). · It must be 2 to 256 characters in length. · It cannot start with http:// or https://.

Figure 9-21: Create a database



5. Click OK.

9.6.2 Modify database description

You can modify the description of a database for easy management. This topic describes how to modify the description of a database.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane of the Basic Information page, choose Database Management > Databases.
4. On the Databases page, click the icon in the Actions column corresponding to the database that you want to modify and choose Change from the shortcut menu.
5. In the Change Database dialog box that appears, set Database Description.

   The database description must meet the following requirements:

   - It must start with a letter and cannot start with http:// or https://.
   - It can contain letters, digits, underscores (_), and hyphens (-).
   - It must be 2 to 256 characters in length.
6. Click OK.

9.6.3 Delete a database

You can delete out-of-date databases in the RDS console.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the Basic Information page.
3. In the left-side navigation pane, choose Database Management > Databases. The Databases page is displayed.
4. Click the icon in the Actions column corresponding to the database that you want to delete, and choose Delete from the shortcut menu. In the Delete Database message that appears, click OK.

9.7 Access mode

RDS supports two access modes: `Standard Mode` and `Safe Mode`. This topic describes the differences between the two access modes and their configuration methods.

Prerequisites

The network type of the instance is Classic Network.

Context

`Standard Mode` **and** `Safe Mode` **have the following differences:**

- **Standard mode: RDS uses Server Load Balancer (SLB) to eliminate the impact of database engine HA switching on the application layer. This mode shortens the response time, but slightly increases the probability of transient disconnections and disables SQL interception.**
- **Safe Mode: This mode prevents 90% of transient disconnections and intercepts SQL injection attacks based on semantic analysis. However, it increases the response time by over 20%.**

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of the instance.**
3. **On the Basic Information page, click Switch Connection Mode.**

Note:
When the access mode change is in progress, `Status` **of the instance changes to** `Switching the access mode`**. When** `Status` **changes to** `Running`**, the access mode is changed.**

9.8 Backup and recovery

9.8.1 RDS data backup

9.8.1.1 Automatic backup

RDS automatic backup supports full physical backups. ApsaraDB for RDS automatically backs up data based on pre-configured policies. This topic describes how to configure a policy for automatic backup.

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of the instance.**
3. **In the left-side navigation pane of the Basic Information page, choose Backup and Restore > Backups.**
4. **On the Backups page, click the Backup Settings tab and click Settings. On the Backup Configurations for Instance page that appears, configure the parameters as prompted.**

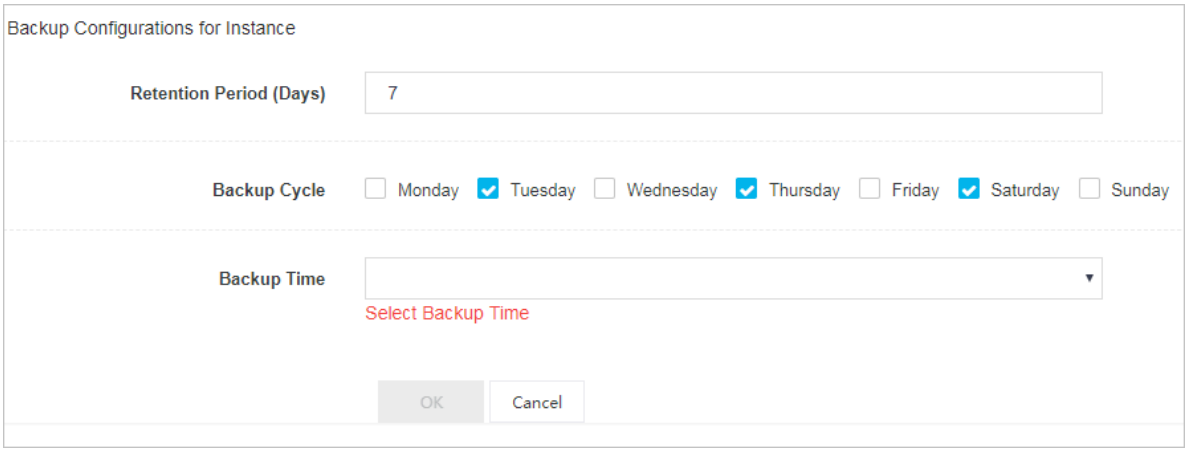
   Table 9-23: Backup policy configuration parameters **describes the parameter configurations.**

   Table 9-23: Backup policy configuration parameters

Parameter	Description
Retention Period (Days)	The number of days, for which backup files are retained. Default value: 7 days. Value range: 1 to 30 days.
Backup Cycle	The day or multiple days in a week, when data is backed up.

Parameter	Description
Backup Time	The specified hour of a day, when data is backed up.

Figure 9-22: Configure a backup policy



5. Click OK.

9.8.1.2 Manual backup

Manual backup supports both full physical backups and full logical backups. This topic describes how to manually back up RDS data.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of an instance.
3. On the Basic Information page, click Back Up Instance. On the Back Up Instance page that appears, set `Backup Mode` and `Backup Type`.

   Instance backup parameters describes the parameter configurations.

   Table 9-24: Instance backup parameters

Parameter	Value	Description
Backup Mode	Physical Backup	This mode dumps the physical files of the RDS database, such as data files, control files, and log files. In case the database fails, these files can be used to restore data.

Parameter	Value	Description
	Logical Backup	This mode stores all schema definition statements and data insertion statements of the RDS database . You can execute these SQL statements to restore data. A database that is exactly the same as the original database is created.
Backup Type	Automatic Backup	This type automatically backs up data based on the preconfigured backup policy.
	Full Backup	This type backs up all files in the database.

Figure 9-23: Configure manual backup

4. Click OK.

9.8.2 RDS data recovery

9.8.2.1 Clone an instance

A cloned instance is a new instance with the same content as the primary instance, including data and settings. This feature allows you to restore data of the primary instance or create multiple instances that are the same as the primary instance.

Prerequisites

To clone an instance, make sure that the primary instance meets the following conditions:

- The instance is in the running state.
- It does not have an ongoing migration task.
- Data backup and log backup are enabled.
- To clone an instance from a backup set, ensure that the primary instance has at least one completed backup set.

Context

You can specify a backup set or any point in time within the backup retention period to clone an instance.

Note:

- **A cloned instance only copies the data of the primary instance. It does not copy the contents of the read-only or disaster recovery instances belonging to the primary instance. The copied data includes database information, account information, and instance configurations such as whitelist configurations, backup configurations, parameter configurations, and alarm threshold configurations.**
- **The database type of a cloned instance must be the same as that of the primary instance. Other settings, such as the instance series, zone, network type, instance specifications, and storage space, can be different. If you want to clone an instance to restore the data of a primary instance, we recommend that you select an instance type and storage space of higher specifications than those of the primary instance to speed up the data recovery process.**
- **The account type of a cloned instance must be the same as that of the primary instance. The account password of the cloned instance can be modified.**

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane of the Basic Information page, choose Backup and Restore > Backups.**
4. **On the Backups page, click the Backups tab.**
5. **On the Backups tab that appears, click Clone Instance. On the Clone Instance page that appears, configure the parameters as prompted.**

   Table 9-25: Instance cloning parameters **describes the parameter configurations.**

   Table 9-25: Instance cloning parameters

Parameter	Description
Restore Mode	The data restoration mode of the primary instance. Valid values: · By Time · By Backup Set

Parameter	Description
Restore Point	The time by which the data in the primary instance is to be restored when the restore mode is `By Time`.
Backup Set	The backup set by which the data in the primary instance is to be restored when the restore mode is `By Backup Set`.
Specifications	The specifications of the cloned instance.
Storage	The storage capacity of the cloned instance.

6. Click Create.

9.8.3 Binary log (binlog)

This topic describes how to check and download the binlogs of an RDS instance.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the Basic Information page, choose Backup and Restore > BinLogs.
4. On the BinLogs page, select a time range and click Search to search for binlogs that are generated within the selected time range.
5. To download a binlog, click the icon in the Actions column and choose Download from the shortcut menu.

9.9 Security

9.9.1 Configure a whitelist

To guarantee database security and reliability, you must modify the whitelist of an RDS instance before you enable the instance. You must add the IP addresses or CIDR blocks that are used for database access to the whitelist of the RDS instance.

Context

The system creates a default whitelist group for each instance. This whitelist group can be modified or cleared, but cannot be deleted. For each newly created RDS instance, IP address 0.0.0.0/0 is added to the default whitelist group by default.

0.0.0.0/0 allows all IP addresses to access the instance, which greatly reduces database security. Delete 0.0.0.0/0 from the whitelist.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane of the Basic Information page, choose Security Control > Whitelist Settings.
4. You can use the following methods to add an IP address or CIDR block:
   - Add an IP address or CIDR block to the default whitelist group.

   a) Click the icon corresponding to the default whitelist group, and add an IP address or CIDR block.

   b) Click OK.

   - Create a whitelist group and add an IP address or CIDR block to the group.

   a) Click Create Whitelist Group. In the Create Whitelist Group dialog box that appears, set Group Name and IP Addresses.

   b) Click OK.

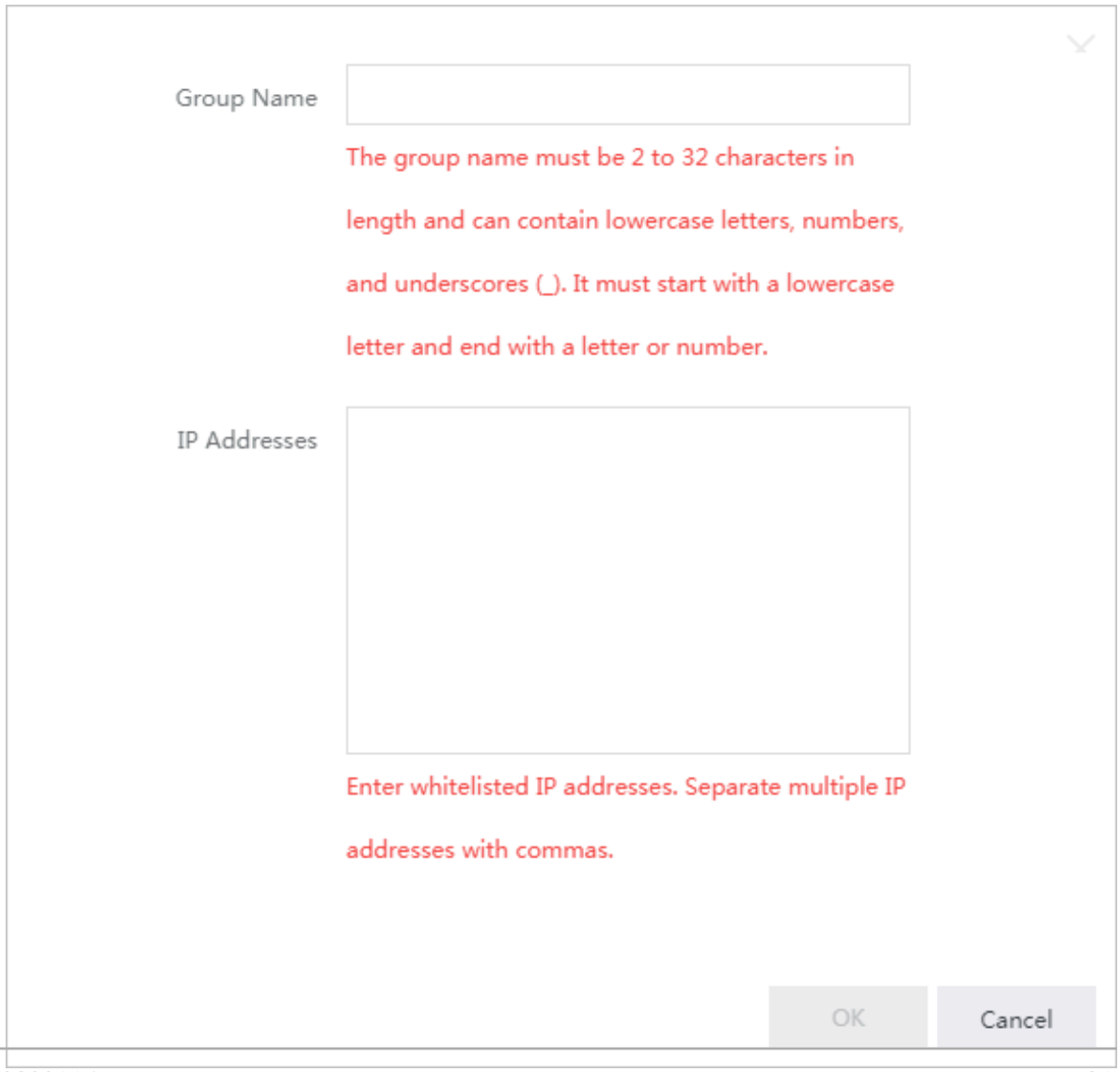
   Table 9-26: Whitelist configuration parameters describes the parameter configurations.

Table 9-26: Whitelist configuration parameters

Parameter	Description
Group Name	The name of the new whitelist group. The naming conventions are as follows: · It must start with a lowercase letter and end with a lowercase letter or digit. · It can contain lowercase letters, digits, and underscores (_). · It must be 2 to 32 characters in length. Note: You cannot change the name of an existing whitelist group.

Parameter	Description
IP Addresses	The IP addresses or CIDR blocks that are allowed to access the RDS instance. **Note:** · If you enter a CIDR block, such as 10.10.10.0/24, any IP addresses in the 10.10.10.X format can access the RDS instance. · If you enter multiple IP addresses, use commas (,) to separate them. Example: 192.168.0.1,172.16.213.9 · 127.0.0.1 indicates that no IP addresses are allowed to access the RDS instance. · 0.0.0.0/0 indicates that all IP addresses are allowed to access the RDS instance.

Figure 9-24: Create a whitelist group



Notice:

The proper use of the whitelist can improve the access security of your RDS instance. We recommend that you maintain the whitelist on a regular basis. After you configure the whitelist, you can perform the following operations:

- **Click the icon to modify the whitelist group.**
- **Click the icon to clear the default whitelist group or delete a custom whitelist group.**

9.9.2 Audit logs

You can query the SQL logs, operation logs, and error logs of an instance in the RDS console to locate and analyze faults.

Context

Note:

Audit logs are stored for seven days, after which they are automatically cleared.

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane of the Basic Information page, choose Security Control > Audit Logs.**
4. **On the Audit Logs page, click the SQL Log, Error Log, Operation Log, or Log Files tab.**

   Table 9-27: Differences among various types of audit logs **describes the differences among SQL logs, error logs, and operation logs.**

Table 9-27: Differences among various types of audit logs

Log type	Description
SQL log	It records the modification operations on all databases.
Error log	It records the SQL statements that failed to be executed on databases in the past month.
Operation log	It records all operations performed in the console.

5. Select a time range and click OK.

What's next

On the Operation Log tab, you can click Export to export operation logs. On the SQL Log tab, you can click Export Log as Audit File to generate log files. Generated log files can be exported from the Log Files tab. Exported files can be used for offline analysis.

9.9.3 Configure SSL encryption

To enhance link security, you can enable Secure Sockets Layer (SSL) encryption and install SSL CA certificates on the necessary application services.

Context

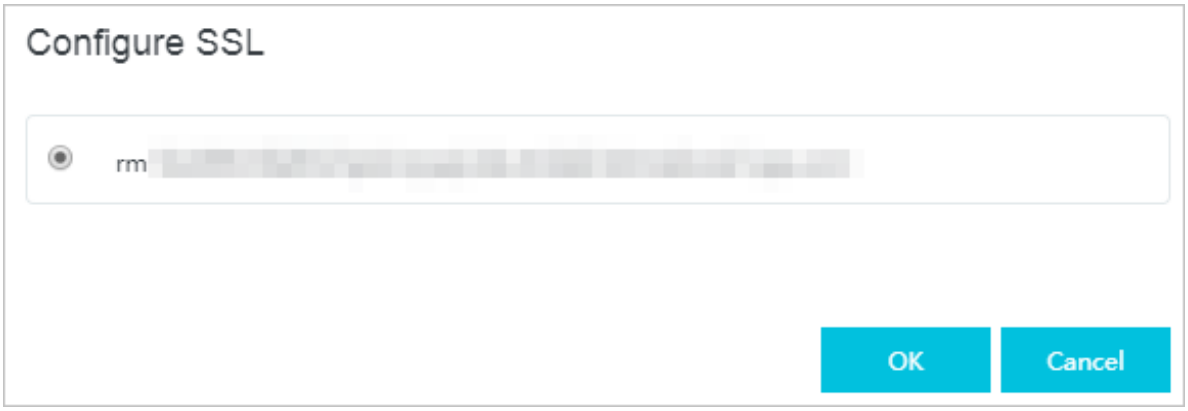
Notice:

- Use caution when enabling SSL encryption, because it cannot be disabled after it is enabled.
- SSL is used on the transport layer to encrypt network connections. It not only increases the security and integrity of communication data, but also increases the response time for network connection.

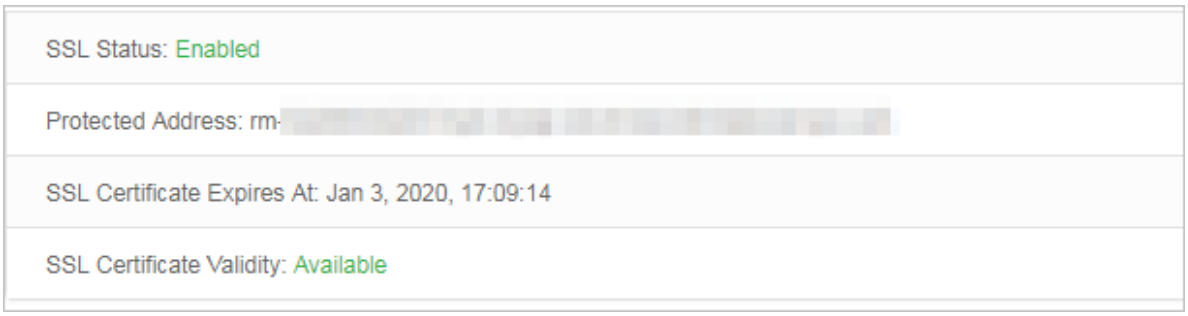
Procedure

1. *Log on to the RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane of the Basic Information page, choose Security Control > SSL Encryption Settings. You can view the SSL details of the instance.
4. Click Enable SSL.

5. **In the Configure SSL dialog box that appears, select the instance ID.**



6. **Click OK to enable SSL encryption, as shown in the following figure.**



9.9.4 Download SSL CA certificates

To enhance link security, you can enable SSL encryption and install SSL CA certificates on the necessary application services.

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane of the Basic Information page, choose Security Control > SSL Encryption Settings. You can view the SSL details of the instance.**

4. Click Download Certificate.

   The downloaded package includes three files:

   - p7b file: used to import CA certificates to the Windows system.
   - PEM file: used to import CA certificates to other operating systems or applications.
   - JKS file: stores truststore certificates in Java. The password is apsaradb. It is used to import the CA certificate chain to Java programs.

    **Note:**

   When the JKS file is used in Java, you must modify the default JDK security configuration in JDK7 and JDK8. Open the `/jre/lib/security/java.security` file on the machine where the database that needs SSL access resides, and modify the following configurations:

   ```
   jdk.tls.disabledAlgorithms=SSLv3, RC4, DH keySize < 224
   jdk.certpath.disabledAlgorithms=MD2, RSA keySize < 1024
   ```

   If you do not modify the JDK security configuration, the following error will be reported. Typically, other similar errors are also caused by Java security configurations.

   ```
   javax.net.ssl.SSLHandshakeException: DHPublicKey does not comply to
    algorithm constraints
   ```

9.9.5 Configure transparent data encryption

Transparent data encryption (TDE) can be used to perform real-time I/O encryption and decryption on data files. Data is encrypted before being written to disks, and decrypted before being read from disks to the memory. TDE does not increase the size of data files. You can use TDE without making changes to applications.

Prerequisites

Before using TDE, make sure that your instance and account meet the following requirements:

- The instance version is ApsaraDB RDS for MySQL 5.6 or 5.7.
- You must log on using an Apsara Stack tenant account (not a RAM user account) to view and modify TDE configurations.

- Before enabling TDE, you must activate Key Management Service (KMS). If you have not activated KMS, you can activate it as instructed when enabling TDE.

Context

> **Note:**
>
> - After TDE is enabled, it cannot be disabled.
> - Keys produced and managed by KMS are used for encryption. RDS does not provide the keys and certificates required for encryption. After enabling TDE , you must decrypt the data through RDS if you want to restore the data to the local server.
> - When TDE is enabled, CPU utilization will significantly increase.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane of the Basic Information page, choose Security Control > TDE Settings. You can view the TDE details of the instance.
4. On the TDE Settings page, click Enable TDE. In the TDE Settings message that appears, click OK.
5. Log on to the database and run the following command to encrypt the relevant tables:

   ```
   alter table <tablename> engine=innodb,block_format=encrypted;
   ```

   Run the following command to decrypt a table that is encrypted with TDE:

   ```
   alter table <tablename> engine=innodb,block_format=default;
   ```

9.10 Read-only instances

9.10.1 Overview

ApsaraDB RDS for MySQL, ApsaraDB RDS for PostgreSQL, and ApsaraDB RDS for PPAS allow you to create read-only instances. In scenarios where there are a few write requests but a large number of read requests, you can create read-only instances to alleviate database access loads on the primary instance. This topic describes the features and limits of read-only instances.

Prerequisites

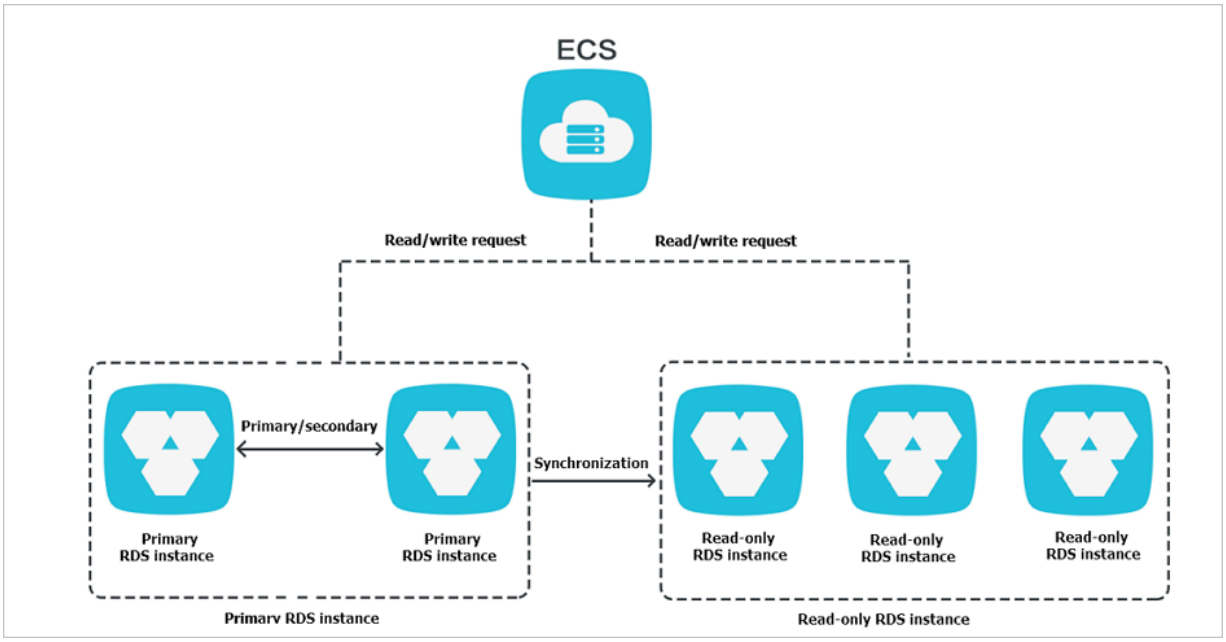
The instance edition must be one of the following editions:

- **MySQL 5.6 or 5.7**
- **PostgreSQL 10.0**
- **PPAS 9.3**

Introduction

To achieve elastic scaling of database read capabilities and distribute database access loads, you can create one or more read-only instances in a region. In this way, large amounts of data can be read from the database and the application throughput can be increased.

A read-only instance with a single physical node and no backup node must be in the same region as the primary instance but does not have to be in the same zone as the primary instance. The following figure shows the topology of a read-only instance.



Read-only instances have the following features:

- **Specifications of a read-only instance can be different from those of the primary instance and can be changed at any time, which facilitates elastic scaling.**
- **Read-only instances do not require account or database maintenance. Account and database information is synchronized from the primary instance.**
- **The whitelists of read-only instances can be configured independently.**

- System performance monitoring is provided.

  ApsaraDB for RDS provides up to 20 system performance monitoring views , including those for disk capacity, IOPS, connections, CPU utilization, and network traffic. You can view the load of instances with ease.
- Optimization recommendations are provided: ApsaraDB for RDS provides a variety of optimization recommendations, such as storage engine check, primary key check, large table check, and check for excessive indexes and missing indexes. You can optimize your databases based on the optimization recommendations and specific applications.

9.10.2 Create a read-only instance

This topic describes how to create a read-only instance based on your business requirements.

Context

Note:

Read-only instances have the following functional limits:

- A maximum of five read-only instances can be created for a primary instance.
- Backup settings and temporary backup are not supported.
- Instance recovery is not supported.
- Data migration to read-only instances is not supported.
- Database creation and deletion are not supported.
- Account creation, deletion, authorization, and password changes are not supported.
- After a read-only instance is created, the primary instance cannot support data recovery by directly overwriting instances with backup sets.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane of the Basic Information page, choose Performance Optimization > Read-Only Instance.

4. **On the Read-Only Instance page, click Create Read-Only Instance. On the Create Read-Only Instance page that appears, configure the parameters as prompted.**

   Table 9-28: Read-only instance creation parameters **describes the parameter configurations.**

   Table 9-28: Read-only instance creation parameters

Parameter	Description
Destination Region	The region to which the read-only instance belongs. This parameter must be the same as that of the primary instance.
Instance Specification	The specifications of the read-only instance. The specifications can be different from those of the primary instance. The specifications of a read-only instance can be modified at any time to facilitate flexible upgrade and downgrade.
Storage Capacity (GB)	The storage space of the read-only instance. To guarantee sufficient I/O throughput for data synchronization, we recommend that read-only instances be allocated at least as much memory as their primary instance.
Network Type	The network type of the read-only instance. You can choose from the following network types: • `Classic Network` • `VPC`: If you use a VPC, we recommend that you choose the same VPC and VSwitch as those of the primary instance.

5. **Click Create.**

9.10.3 View read-only instance details

9.10.3.1 View instance details through a read-only instance

You can go to the read-only instance management page from the RDS Instances page or the Read-Only Instance page of the primary instance. Read-only instances are managed in the same way as ordinary instances. The page shows the management operations that can be performed. This topic describes how to go to the read-only instance management page from the RDS Instances page.

Procedure

1. *Log on to the RDS console*.

2. **On the RDS Instances page, click the ID of the read-only instance. The Basic Information page that appears allows you to manage the read-only instance.**

   In the instance list, `Instance Type` **of read-only instances is displayed as** `Read-Only Instance`**, as shown in** *Figure 9-25: View read-only instances*.

   Figure 9-25: View read-only instances

9.10.3.2 View instance details through the primary instance

You can go to the read-only instance management page from the RDS Instances page or the Read-Only Instance page of the primary instance. Read-only instances are managed in the same way as ordinary instances. The page shows the management operations that can be performed. This topic describes how to go to the read-only instance management page from the Read-Only Instance page of the primary instance.

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of the instance.**
3. **In the left-side navigation pane of the Basic Information page, choose Performance Optimization > Read-Only Instance.**
4. **On the Read-Only Instance page, click the ID of a read-only instance. The Basic Information page that appears allows you to manage the read-only instance.**

9.11 Read/write splitting

9.11.1 Overview

This topic describes the principles and benefits of read/write splitting and how it can be used.

The primary RDS instance and all its read-only RDS instances each have their own independent connection address. The connection addresses are configured by an application for data read/write splitting.

The read/write splitting function provides an extra read/write splitting address . This address represents the primary instance and all its read-only instances.

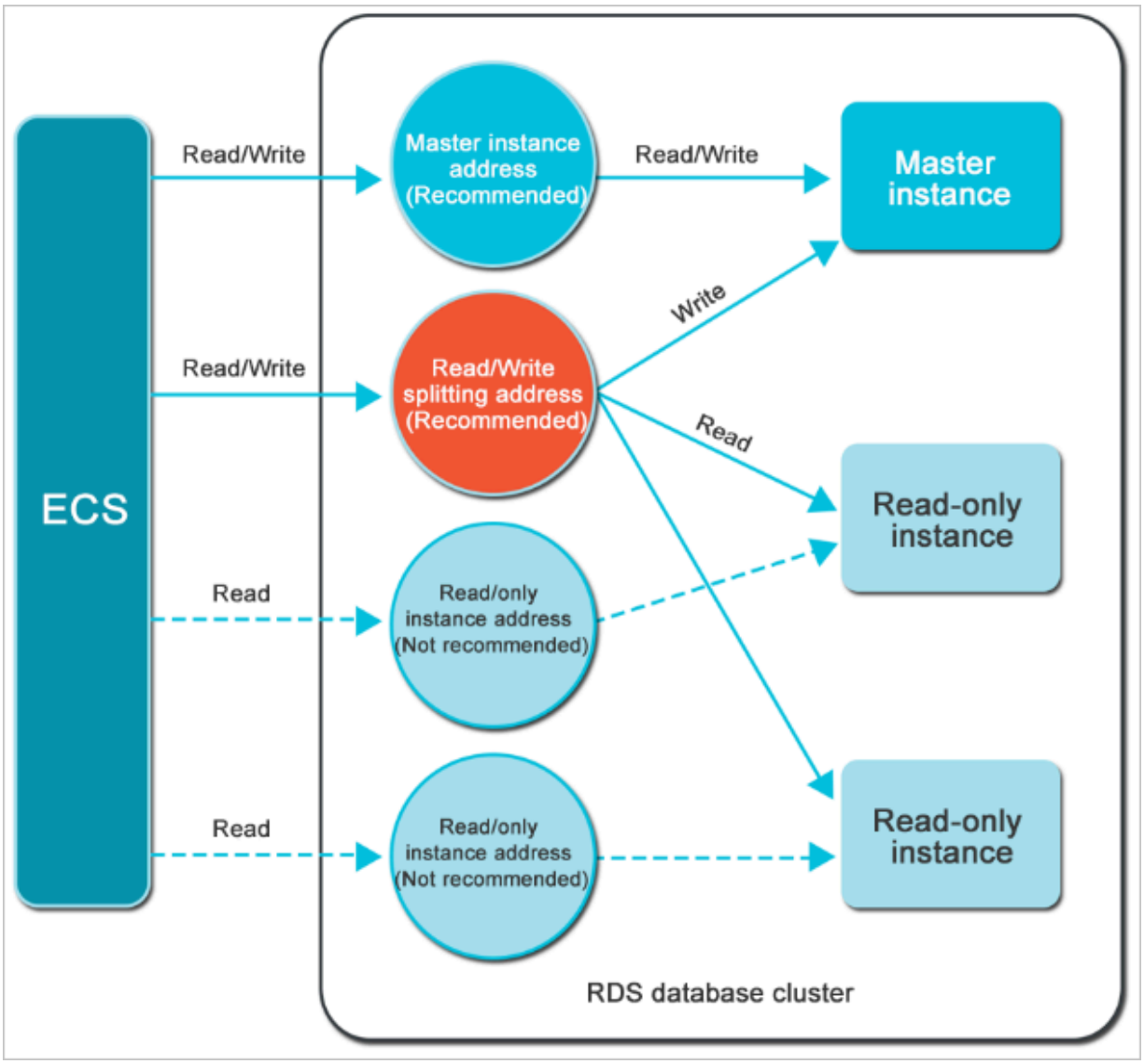
Requests sent to this address are automatically distributed based on read/write splitting. An application only needs to connect to the read/write splitting address for both read and write requests. Write requests are automatically routed to the primary instance, and read requests are routed to the read-only instances based on their weights. You can scale out the processing capability of the system by adding more read-only instances. No application changes are required.

Read/write splitting is only supported in ApsaraDB RDS for MySQL 5.6 or 5.7. When read/write splitting is enabled, there will be three types of addresses for the instances:

- **Connection address of the primary instance**
- **Connection addresses of the read-only instances**
- **Connection address of read/write splitting**

Principle of read/write splitting **shows how an application uses different types of connection addresses to access databases.**

Figure 9-26: Principle of read/write splitting



Read/write splitting has the following benefits:

- **Facilitates maintenance with a single read/write splitting address.**

  In the current read-only account mode, the primary instance and all its read-only instances each have their own independent connection address. You must configure each of the addresses in your application to have write requests sent to the primary instance and read requests sent to the read-only instances.

  RDS read/write splitting provides a read/write splitting address in addition to the existing instance connection addresses. After an application is connected to

the read/write splitting address, it can perform read and write operations on the corresponding primary and read-only instances. The forwarding logic of read and write statements is transparent to the user. This reduces the maintenance cost.

- Improves RDS performance with high-security link support.

  For users who build a proxy layer to implement read/write splitting on the cloud, data must pass through multiple components for statement parsing and forwarding before it reaches a database. This significantly increases the response latency. RDS read/write splitting is built on the existing high-security link and does not require any additional components. This greatly reduces the latency and improves the processing efficiency.
- Applies to various use cases with configurable weights and thresholds.

  RDS read/write splitting allows you to set read request weights of the primary and read-only instances, as well as the latency threshold of read-only instances.
- Enhances database availability with instance health check.

  RDS read/write splitting automatically performs health checks on all instances in the distribution system. If an instance fails a health check or its latency exceeds the threshold, RDS automatically removes the instance from the distribution system, making it unavailable for read request allocation. RDS then allocates read and write requests to the remaining instances based on predefined weights. This ensures normal application access in case of read-only instance failures. After the instance is restored, RDS automatically reclaims it into the request distribution system.

  Note:
  We recommend that you create at least two read-only instances for the primary instance when using read/write splitting. This ensures normal database access in case of a single point of failure.

9.11.2 Enable read/write splitting

In scenarios where there are a few write requests but a large number of read requests, you can enable read/write splitting to distribute the database read load on the primary instance. This topic describes how to enable read/write splitting.

Prerequisites

- The instance is a high-availability ApsaraDB RDS for MySQL 5.6 or 5.7 primary instance.
- A read-only instance has been created under the primary instance. If there are no read-only instances, you must create one. For more information about how to create a read-only instance, see *Create a read-only instance*.
- The primary instance has been switched to safe mode.

Context

When you enable read/write splitting for the first time, the system automatically upgrades the back-end control system of the primary instance and all associated read-only instances to the latest version to guarantee normal service operations. When the function is enabled, the primary and read-only instances automatically restart once. During the restart process, the primary instance is subject to a transient disconnection of 30s or less, and the read-only instances are inaccessible. To avoid service impacts from transient disconnections, we recommend that you enable read/write splitting during off-peak hours and make sure that your application can be automatically reconnected.

 **Note:**

- The following commands or functions cannot be forwarded to a read-only instance:
  - The `stmt prepare sql` command is automatically executed on the primary instance.
  - The `stmt prepare` command cannot be forwarded to a read-only instance before the execution of the `stmt close` command.
  - `set global`, `set user`, and `set once` are automatically executed on the primary instance.
- The execution result is random for the following commands:

  `show processlist`, `show master status`, and `com_process_info` return results based on the instance that is connected to during command execution.
- All transactions are routed to the primary database.
- Read/write splitting does not guarantee consistency of non-transactional reads. If you require such consistency, add a hint to route query requests to the primary database or encapsulate the query requests within transactions.

- **The following commands or functions are not supported:**
  - **SSL encryption**
  - **Compression protocols**
  - **com_dump_table and com_change_user protocols**
  - `kill connection [query]`
  - `change user`

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane of the Basic Information page, choose Performance Optimization > Read/Write Splitting.**
4. **On the Read/Write Splitting page, click Enable. In the Configure Read/Write Splitting dialog box that appears, configure the parameters as prompted.**

   Table 9-29: Read/write splitting parameters **describes the parameter configurations.**

   Table 9-29: Read/write splitting parameters

Parameter	Description
SLB Type	The read/write splitting address. Valid values: Private Address (VPC) and Public URL.
Latency Threshold	The maximum latency allowed when read-only instances synchronize data from the primary instance. The value range is 1 to 7,200 seconds. If the latency of a read-only instance exceeds this threshold, read requests are not forwarded to this instance regardless of the instance weight. Read-only instances may experience latency depending on the execution status of SQL statements. We recommend that you do not set the latency threshold to less than 30s.

Parameter	Description
Weights of Read Requests	The read request weight of each instance. An instance with a higher weight is allocated more read requests. For example, a read-write splitting address has a primary instance and three read-only instances. The read weights of the instances are 0, 100, 200, and 200. This means that the primary instance does not process any read requests while the three read-only requests are allocated read requests in a ratio of 1:2:2. You can customize read request weights or allow the system to distribute read request weights automatically. · Default: The system automatically distributes weights to instances based on their specifications. Newly created read-only instances under the primary instance are automatically added to the read/write splitting link based on a preset weight. For more information, see *Rules of system weight distribution*. · Customized: You can customize the read request weight of an instance. The weight can be from 0 to 10,000 and must be a multiple of 100. In this mode, the read request weights of newly created read-only instances under the primary instance are set to 0 by default. You must manually set this parameter to distribute requests to the read-only instances.

5. Click OK.

9.11.3 Modify the latency threshold and weights of read requests

After read/write splitting is enabled, you can modify the latency threshold and weights of read requests.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane of the Basic Information page, choose Performance Optimization > Read/Write Splitting.

4. **On the Read/Write Splitting page, click Enable. In the Configure Read/Write Splitting dialog box that appears, configure the parameters as prompted.**

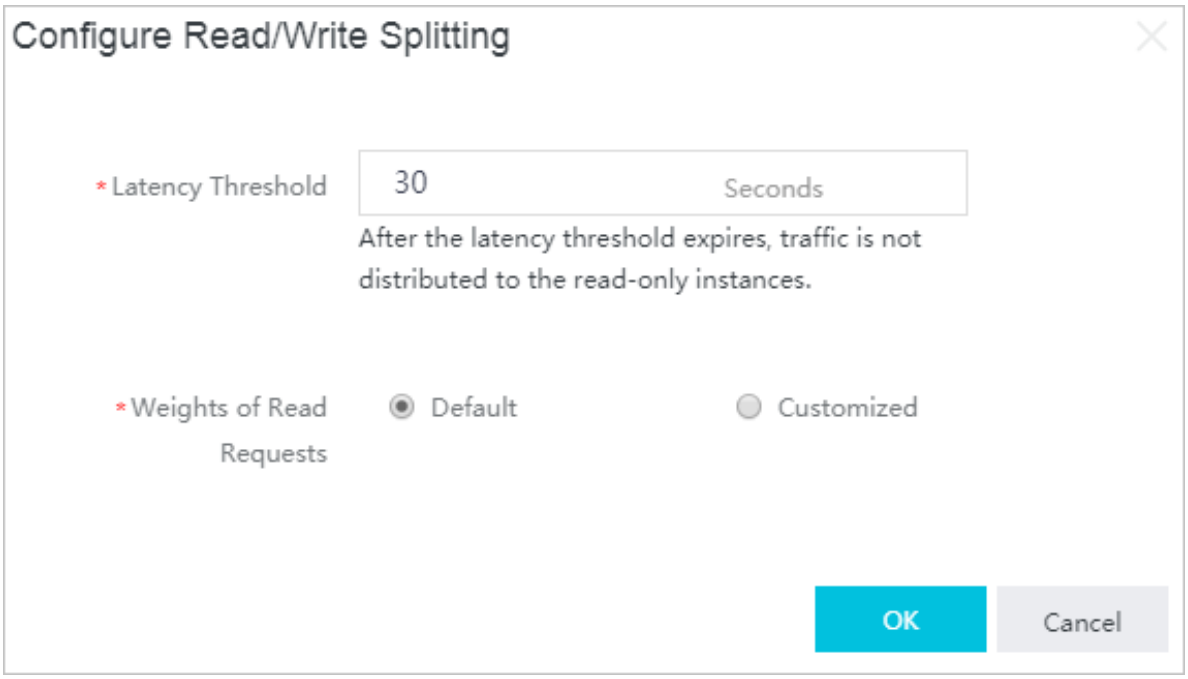
   Table 9-30: Read/write splitting parameters **describes the parameter configurations.**

   Table 9-30: Read/write splitting parameters

Parameter	Description
Latency Threshold	The maximum latency allowed when read-only instances synchronize data from the primary instance. The value range is 1 to 7,200 seconds. If the latency of a read-only instance exceeds this threshold, read requests are not forwarded to this instance regardless of the instance weight. Read-only instances may experience latency depending on the execution status of SQL statements. We recommend that you do not set the latency threshold to less than 30s.

Parameter	Description
Weights of Read Requests	The read request weight of each instance. An instance with a higher weight is allocated more read requests. For example, a read/write splitting address has a primary instance and three read-only instances. The read weights of the instances are 0, 100, 200, and 200. This means that the primary instance does not process any read requests while the three read-only requests are allocated read requests in a ratio of 1:2:2. You can customize read request weights or allow the system to distribute read request weights automatically. · Default: The system automatically distributes weights to instances based on their specifications. Newly created read-only instances under the primary instance are automatically added to the read/write splitting link based on a preset weight. For more information, see *Rules of system weight distribution*. · Customized: You can customize the read request weight of an instance. The weight can be from 0 to 10,000 and must be a multiple of 100. In this mode, the read request weights of newly created read-only instances under the primary instance are set to 0 by default. You must manually set this parameter to distribute requests to the read-only instances.

Figure 9-27: Modify read/write splitting parameters



5. Click OK.

9.11.4 Disable read/write splitting

You can disable read/write splitting if it is no longer needed. This topic describes how to disable read/write splitting.

Context

> **Note:**
>
> - **Read/write splitting can be used only when at least one read-only instance is available. Therefore, to be able to delete the last read-only instance, you must disable read/write splitting.**
> - **After read/write splitting is disabled, your application can no longer connect to the read/write splitting address. Make sure that your database connection configuration does not include this connection address.**

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of the instance.**
3. **In the left-side navigation pane of the Basic Information page, choose Performance Optimization > Read/Write Splitting.**
4. **On the Read/Write Splitting page, click Disable.**
5. **Click OK.**

9.11.5 Monitor read/write splitting performance

You can view the read/write splitting performance on the monitoring page of the RDS console.

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane of the Basic Information page, choose System Resource Monitoring > Database Performance.**
4. **On the Database Performance page, click the QPS/TPS tab to view Average Transactions per Second (TPS) / Average SQL Queries per Second (QPS). You can view the number of reads and writes of all databases involved in read/write splitting, including the primary database and read-only databases.**

9.11.6 Rules of system weight distribution

RDS automatically distributes weights to instances based on their configurations. This topic describes the rules for the read weight distribution system and how to use hints to specify whether an SQL statement is to be sent to the primary instance or read-only instances.

Weight values list

The system automatically configures fixed read weight values for instances, as listed in *Table 9-31: Weight values*.

Table 9-31: Weight values

Specification code	Specification type	Memory	CPU	Weight
rds.mysql.t1.small	Common instance	1 GB	1	100
rds.mysql.s1.small	Common instance	2 GB	1	100
rds.mysql.s2.large	Common instance	4 GB	2	200
rds.mysql.s2.xlarge	Common instance	8 GB	2	200
rds.mysql.s3.large	Common instance	8 GB	4	400
RDS. MySQL. m1.medium	Common instance	16 GB	4	400
rds.mysql.c1.large	Common instance	16 GB	8	800
rds.mysql.c1.xlarge	Common instance	32 GB	8	800
rds.mysql.c2.xlarge	Common instance	64 GB	16	1,600
rds.mysql.c2.xlp2	Common instance	96 GB	16	1,600
mysql.x8.medium. 2	Dedicated instance	16 GB	2	200

Specification code	Specification type	Memory	CPU	Weight
mysql.x8.large. 2	Dedicated instance	32 GB	4	400
mysql.x8.xlarge. 2	Dedicated instance	64 GB	8	800
mysql.x8.2xlarge. 2	Dedicated instance	128 GB	16	1,600
rds.mysql.st.d13	Dedicated host	220 GB	30	3,000

Use hints to specify whether an SQL statement is to be sent to the primary instance or read-only instances

In addition to the weight distribution system for read/write splitting, hints are used as supplementary SQL syntax to force SQL statements to be executed on the primary instance or read-only instances.

The hint formats supported by RDS read/write splitting are as follows:

- `/*FORCE_MASTER*/`**: specifies that the following SQL statements are to be executed on the primary instance.**
- `/*FORCE_SLAVE*/`**: specifies that the following SQL statements are to be executed on the read-only instances.**

For example, after the following statement is prefixed with a hint, the statement is always routed to and executed on the primary instance regardless of preset weight.

```
/*FORCE_MASTER*/ SELECT * FROM table_name;
```

9.12 Performance optimization

9.12.1 Slow SQL statistics

You can use the RDS console to query slow SQL statistics to locate and analyze faults.

Procedure

1. *Log on to the RDS console*.
2. **Click the ID of an instance.**

3. In the left-side navigation pane of the Basic Information page, choose Performance Optimization > Slow SQL Statistics.
4. On the Slow SQL Statistics page, select a time range and click Search.

> **Note:**
> The system does not list slow logs from the past two hours. These logs are contained in the slow_log table of a MySQL database.

9.12.2 Query missing indexes

Based on the SQL statement execution status and performance of your RDS instances, the system prompts you about database tables with missing indexes, and provides you with a statement to add indexes.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane of the Basic Information page, choose Performance Optimization > Indexes. This page allows you to query all missing indexes.

9.13 Monitor system resources

The RDS console provides a variety of performance metrics for you to monitor the status of your instances.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane of the Basic Information page, click System Resource Monitoring.

4. **Select the monitored data that you want to view, such as System Resources, Database Performance, InnoDB Engine, and MyISAM Engine.** *Table 9-32: Metrics* **lists the metrics on each page.**

   Table 9-32: Metrics

Page	Metric	Description
System Resources	Disk Space	The disk space usage of the instance, such as overall usage of the disk space, data space, log space, temporary file space, and system file space. Unit: MB
	IOPS	The number of I/O requests for the instance per second.
	CPU Utilization	The CPU utilization of the instance (excluding the CPU resources used by the operating system).
	Network Traffic	The inbound and outbound traffic of the instance. Unit: Kbit/s
Database Performance	QPS/TPS	The number of SQL statements executed and transactions processed per second.
	Temporary Tables	The number of temporary tables that are automatically created on the hard disk when the database executes SQL statements.
	COMDML	The number of SQL statements executed by the database per second. The statements counted in this parameter include INSERT, DELETE, INSERT_SELECT, REPLACE, REPLACE_SE LECT, SELECT, and UPDATE.
	ROWDML	The number of operations performed on InnoDB per second, such as the number of physical writes to the log file, and the numbers of InnoDB table rows that are read, updated, deleted, and inserted per second.
InnoDB Engine	InnoDB Buffer Pool	The read hit rate, utilization, and dirty data block percentage of the InnoDB buffer pool.

Page	Metric	Description
	InnoDB Read /Write	The volume of InnoDB data that is read and written per second. Unit: KB
	InnoDB Reads and Writes	The number of InnoDB reads and writes per second.
	InnoDB Log	The numbers of InnoDB physical writes to the log file, log write requests, and FSYNC writes to the log file.
MyISAM Engine	MyISAM Key Buffer	The read hit rate, write hit rate, and usage of MyISAM key buffers per second.
	MyISAM Reads and Writes	The numbers of MyISAM reads and writes from and to the buffer pool and hard disk per second.

9.14 Migrate a local database to RDS

9.14.1 Compress data

RDS for MySQL 5.6 allows you to compress data by using the TokuDB storage engine. This topic describes how to compress data.

Context

Extensive tests show that the data volume can be reduced by 80% to 90% after data tables are converted from the InnoDB storage engine to the TokuDB storage engine . For example, 2 TB of data can be compressed to 400 GB or less by using TokuDB . In addition to data compression, TokuDB supports transaction and online DDL operations. It is also compatible with MyISAM or InnoDB applications.

Note:

- **The TokuDB storage engine does not support foreign keys.**
- **The TokuDB storage engine is not suitable for scenarios where large amounts of data is read.**

Procedure

1. **Run the following command to check the MySQL version:**

```
SELECT version();
```

2. **Run the following command and set `loose_tokudb_buffer_pool_ratio` to indicate the proportion of cache that TokuDB occupies in the shared cache of TokuDB and InnoDB:**

```
select sum(data_length) into @all_size from information_schema.
tables where engine='innodb';
select sum(data_length) into @change_size from information_schema
.tables where engine='innodb' and concat(table_schema, '.',
table_name) in ('XX.XXXX', 'XX.XXXX', 'XX.XXXX');
select round(@change_size/@all_size*100);
```

> **Note:**
> **In the preceding command, `XX.XXXX` indicates the name of the database or table to be transferred to the TokuDB storage engine.**

3. **Restart the instance. For more information, see** *Restart an instance*.
4. **Run the following command to modify the storage engine: You can also log on to DMS to modify the storage engine. For more information, see** *DMS Documentation*.

```
ALTER TABLE XX.XXXX ENGINE=TokuDB
```

> **Note:**
> **In the preceding command, XX.XXXX indicates the name of the database or table to be transferred to the TokuDB storage engine.**

9.14.2 Migrate MySQL data

9.14.2.1 Use DTS to migrate MySQL data

Data Transmission Service (DTS) can migrate data from a local database to an ApsaraDB RDS for MySQL instance without stopping services. This topic describes how to migrate data from a local database with a private IP address to an ApsaraDB RDS for MySQL instance.

Background information

DTS allows you to perform schema migration, full data migration, and incremental data migration on MySQL databases.

- **Schema migration**

  DTS migrates the schema definition of a local database to the destination instance. DTS supports schema migration for the following objects: tables, views , triggers, stored procedures, and storage functions.

- **Full data migration**

  DTS migrates all existing data of objects from a local database to the destination instance. If you also select incremental data migration, non-transaction tables without primary keys are locked during the full data migration process. Data cannot be written to these locked tables, and the locking duration depends on the data volume of the tables. The locks are released only after these tables are migrated. In this way, data consistency is guaranteed.

- **Incremental data migration**

  In incremental data migration, data changes made during the migration are updated to the destination instance. If DDL operations are performed during migration, the schema changes are not migrated to the destination instance.

Restrictions

Migrating data from a local database to an ApsaraDB RDS for MySQL instance is subject to the following restrictions:

- **DDL operations are not supported during migration.**
- **Event migration is not supported in schema migration.**
- **If object name mapping is enabled, other objects dependent on this object may fail to be migrated.**
- **When incremental data migration is selected, binlogging must be enabled and binlog_format must be set to row for the local MySQL instance. If the version of the local MySQL database is 5.6, binlog_row_image of the database must be set to full.**

Prerequisites

You have added a whitelist for the ApsaraDB RDS for MySQL instance. For more information, see *Configure a whitelist*.

Prepare local data

Before the migration, create the migration accounts in the local database and ApsaraDB RDS for MySQL instance. You also need to create the database to be

migrated in the RDS for MySQL instance, and grant the read and write permissions of the database to the migration account. *Table 9-33: Migration types and required permissions* lists the permissions required by the migration accounts of the source and destination instances when different migration types are used.

Table 9-33: Migration types and required permissions

Database type	Schema migration	Full data migration	Incremental data migration
Local database	select	select	· select · replication slave · replication client
RDS for MySQL instance	Read and write permissions	Read and write permissions	Read and write permissions

1. Run the following command to create a migration account in the local database:

   ```
   CREATE USER 'username'@'host' IDENTIFIED BY 'password';
   ```

   Parameter description:

   - username: specifies the name of the account to be created.
   - host: specifies the host from which you log on to the database by using the account. As a local user, you can use `localhost` to log on to the database. To log on from any hosts, you can use wildcard `%`.
   - password: specifies the logon password for the account.

   For example, if you want to create account `William` with password `Changme123` for logging on to the local database from any hosts, run the following command:

   ```
   CREATE USER 'William'@'%' IDENTIFIED BY 'Changme123';
   ```

2. **Grant permissions to the migration account in the local database.** *Table 9-33: Migration types and required permissions* **lists the permissions required for the migration account of the local database.**

```
GRANT privileges ON databasename.tablename TO 'username'@'host' WITH
 GRANT OPTION;
```

Parameter description:

- **privileges: specifies the operation permissions granted to the account, such as SELECT, INSERT, and UPDATE. To grant all permissions to the account, use** `ALL` **.**
- **databasename: specifies the database name. To grant all database permissions to the account, use wildcard** `*`**.**
- **tablename: specifies the table name. To grant all table permissions to the account, use wildcard** `*`**.**
- **username: specifies the name of the account to be granted permissions.**
- **host: specifies the host from which the account is authorized to log on to the database. As a local user, you can use** `localhost` **to log on to the database. To log on from any hosts, you can use wildcard** `%`**.**
- **WITH GRANT OPTION: specifies an optional parameter that enables the account to use the GRANT command.**

For example, if you want to grant all of the database and table permissions to account `William` **and use the account to log on to the local database from any hosts, run the following command:**

```
GRANT ALL ON *. * TO 'William'@'%';
```

Note:
If you want to perform incremental data migration, follow these steps to enable binlogging on the local database and configure this feature correctly.

3. **Run the following command to check whether binlogging has been enabled:**

```
show global variables like "log_bin";
```

If the query result is `log_bin=OFF`**, binlogging is not enabled on the local database. To ensure synchronous migration of the incremental data generated**

during the migration process, modify the following parameters in the *my.cnf* **configuration file:**

```
log_bin=mysql_bin
binlog_format=row
server_id = integer greater than 1
binlog_row_image=full //When the local MySQL version is later than 5.6
, this item must be set.
```

4. **After the parameters are set, run the following commands to restart the MySQL process:**

```
$mysql_dir/bin/mysqladmin -u root -p shutdown
$mysql_dir/bin/safe_mysqld &
```

mysql_dir **is the installation directory of MySQL.**

Procedure

Perform migration after data preparation is completed.

1. **Log on to the DTS console.**
2. **In the left-side navigation pane, click Data Migration. On the page that appears, click Create Migration Task in the upper-right corner. In the Create DTS Instances dialog box that appears, create an instance as prompted.**

   Table 9-34: DTS instance creation parameters **describes the parameter configurations.**

   Table 9-34: DTS instance creation parameters

Parameter	Description
Feature	The feature specified by the system. In this case, the value is Data Migration.
Region	The region where the instance is located.
Instances to Create	The number of instances to be created.

3. **Click OK.**

4. **In the migration task list, find the instance that you created and click Configure Migration Task. On the Create Migration Task page that appears, configure the parameters as prompted.**

   Table 9-35: Parameters for creating a migration task **describes the parameter configurations.**

   Table 9-35: Parameters for creating a migration task

Category	Parameter	Description
N/A	Task Name	DTS automatically generates a name for each task. You can change the default name to an informative one for easy task identification.
Source Database	Instance Type	The type of the source instance. Select User-Created Database with Public IP Address.
	Instance Region	The region where the source instance is located.
	Database Type	The type of the database to be migrated. Select MySQL.
	Hostname or IP Address	The connection address of the database to be migrated.
	Port	The port number of the database to be migrated. The default port number for a MySQL database is 3306.
	Database Account	The account used to log on to the database to be migrated.
	Database Password	The password of the account used to log on to the database to be migrated.
Destination Database	Instance Type	The type of the destination instance. Select RDS Instance.
	Instance Region	The region where the ApsaraDB RDS for MySQL instance is located. Select the same region as that of the source instance.
	RDS Instance ID	The ID of the ApsaraDB RDS for MySQL instance.

Category	Parameter	Description
	Database Account	The account used to log on to the ApsaraDB RDS for MySQL instance.
	Database Password	The password of the account used to log on to the ApsaraDB RDS for MySQL instance.

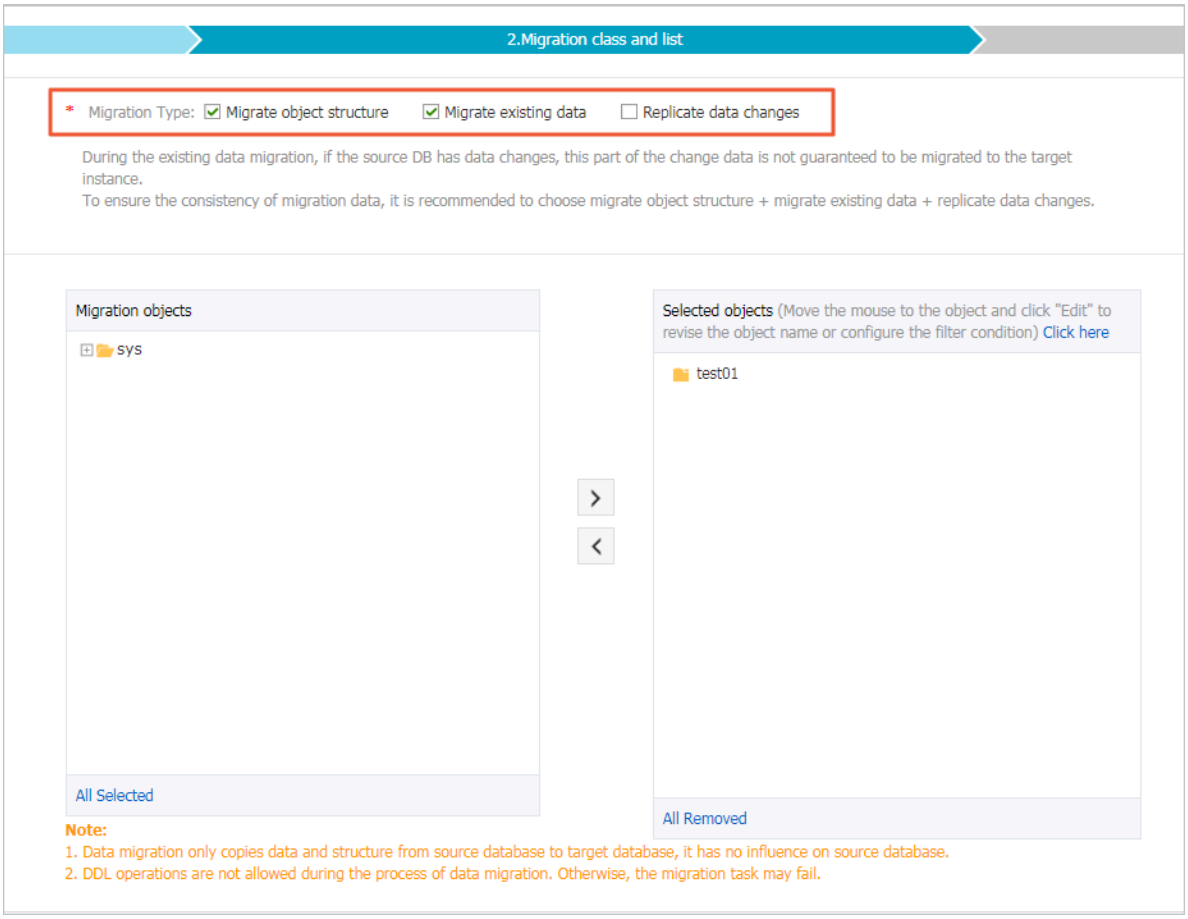
Note:
After configuring the source and destination databases, you can click Test Connectivity to test the connectivity.

5. After configuring the preceding parameters, click Set Whitelist and Next.
6. On the Configure Migration Types and Objects page, select migration types. In the Available section, select the objects to be migrated, and click > to add the selected objects to the Selected section, as shown in *Figure 9-28: Select migration types and objects to be migrated*.

Note:
To modify the name of an object to be migrated in the destination database, move the pointer over the database to be modified in the Selected section. The

Edit button is displayed, as shown in *Figure 9-28: Select migration types and objects to be migrated*.

Figure 9-28: Select migration types and objects to be migrated



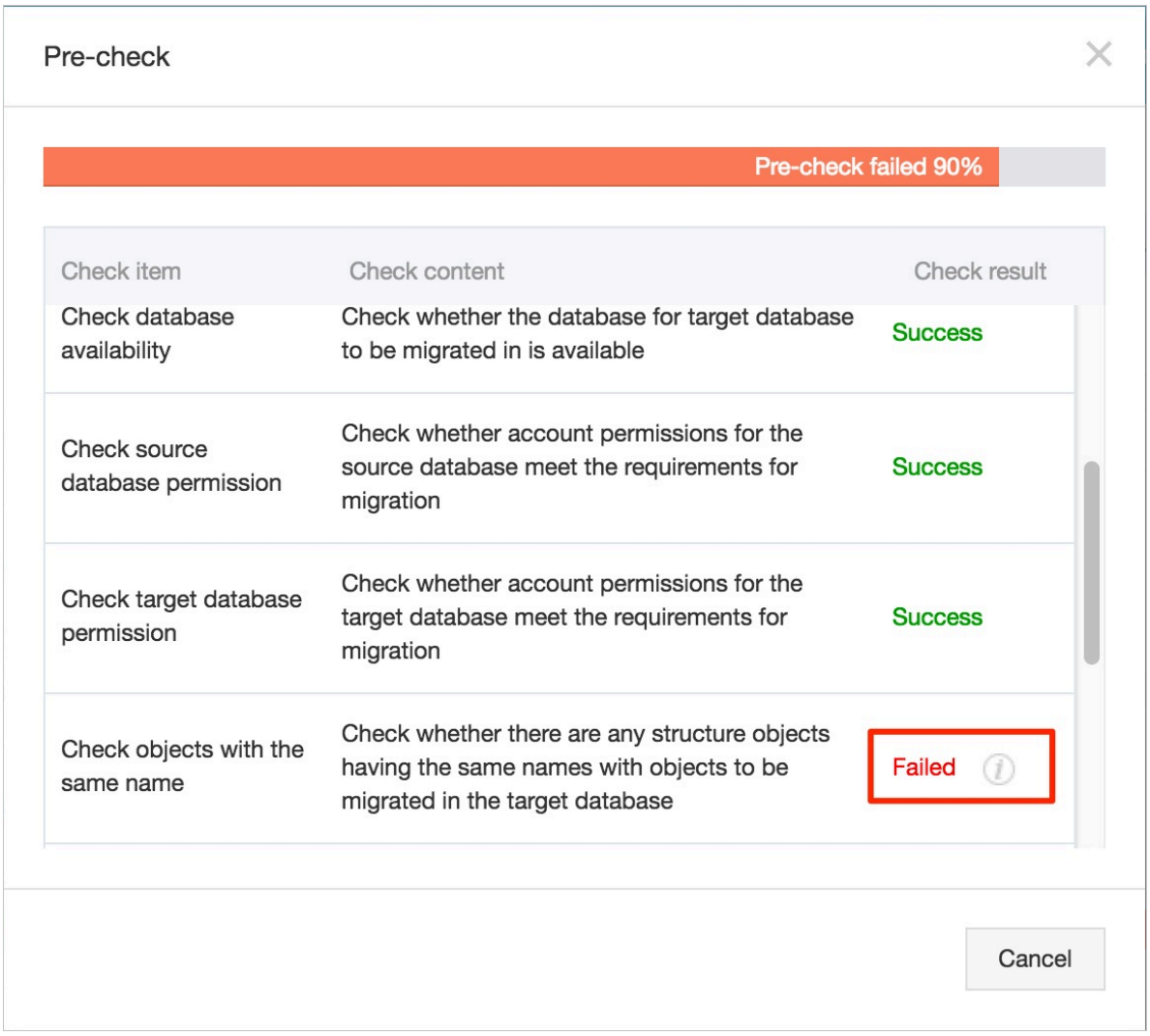
7. **Click Precheck, as shown in *Figure 9-29: Precheck*.**

Note:

- **A precheck is required before you can start the migration task. A migration task can only be started after it passes the precheck.**
- **If the precheck fails, click the info icon corresponding to each failed check item to view the failure details, troubleshoot the faults, and re-run the precheck.**

- **After you rectify the faults, select the task from the task list and restart the precheck.**

Figure 9-29: Precheck



8. **After the precheck succeeds, you can start the migration task. After the task starts, you can check the migration status and progress on the Migration Tasks page.**

Subsequent operations

The migration accounts have been granted the read and write permissions. For security considerations, we recommend that you delete the accounts from the local database and the ApsaraDB RDS for MySQL instance after the data migration.

9.14.2.2 Use mysqldump to migrate MySQL data

This topic describes how to use mysqldump to migrate local data to RDS for MySQL.

Prerequisites

An ECS instance has be activated.

Context

mysqldump is easy to use but has long downtimes. The tool is suitable for scenarios where the amount of data is small or long downtimes are allowed.

RDS for MySQL is fully compatible with the native database service. The procedure to migrate the original database to an RDS for MySQL instance is similar to that of migrating data from one MySQL server to another.

Before you migrate data, create a migration account in the local database, and grant the read and write permissions on the database to the migration account.

Procedure

1. **Run the following command to create a migration account in the local database:**

   ```
   CREATE USER 'username'@'host' IDENTIFIED BY 'password';
   ```

   Parameter description:

   - **username: specifies the name of the account to be created.**
   - **host: specifies the database host to which the account logs on. As a local user, you can use `localhost` to log on to the database. To enable the account to log on to any host, you can use wildcard `%`.**
   - **password: specifies the password that is used to log on to the account.**

   The following example creates an account named `William` with password `Changme123`, which is allowed to log on to the local database from any host.

   ```
   CREATE CREATEUSER'William'@'%' IDENTIFIED BY 'Changme123';
   ```

2. **Run the following command to authorize the migration account of the local database:**

   ```
   GRANT SELECT ON databasename.tablename TO 'username'@'host' WITH
    GRANT OPTION; GRANT REPLICATION SLAVE ON databasename.tablename
   ```

```
TO 'username'@'host' WITH GRANT OPTION;GRANT REPLICATION SLAVE ON
databasename.tablename TO 'username'@'host' WITH GRANT OPTION;
```

Parameter description:

- **privileges: specifies the operation permissions of the account, such as SELECT, INSERT, and UPDATE. To grant all permissions to the account, use `ALL`.**
- **databasename: specifies the database name. To grant all database permissions to the account, use wildcard `*`.**
- **Tablename: specifies the table name. To grant all table permissions to the account, use wildcard `*`.**
- **username: specifies the name of the account to be granted permissions.**
- **host: specifies the host, from which the account is authorized to log on to the database. As a local user, you can use `localhost` to log on to the database. To log on from any host, you can use wildcard `%`.**
- **WITH GRANT OPTION: specifies an optional parameter that enables the account to use the GRANT command.**

In the following command, the account named `William` is granted all database and table permissions, and allowed to log on to the local database from any host:

```
GRANT ALL ON *. * TO 'William'@'%';
```

3. **Use the data export tool of mysqldump to export data from the database as data files.**

Notice:
Do not update data during the data export. This step exports data only. It does not export stored procedures, triggers, or functions.

```
mysqldump -h localIp -u userName -p --opt --default-character-set=utf8
 --hex-blob dbName --skip-triggers > /tmp/dbName.sql
```

Parameter description:

- **localIp: specifies the IP address of the local database server.**
- **userName: specifies the migration account of the local database.**
- **dbName: specifies the name of the database to be migrated.**
- **/tmp/dbName.sql: specifies the name of the backup file.**

4. **Use mysqldump to export stored procedures, triggers, and functions.**

   > **Notice:**
   > **Skip this step if no stored procedures, triggers, or functions are used in the database. When you export stored procedures, triggers, or functions, you must remove the definer to be compatible with RDS.**

   ```
   mysqldump -h localIp -u userName -p --opt --default-character-set=utf8
    --hex-blob dbName -R | sed -e 's/DEFINER[ ]*=[ ]*[^*]*\*/\*/' > /tmp/
   triggerProcedure.sql
   ```

   Parameter description:

   - **localIp: specifies the IP address of the local database server.**
   - **userName: specifies the migration account of the local database.**
   - **dbName: specifies the name of the database to be migrated.**
   - **/tmp/triggerProcedure.sql: specifies the name of the backup file.**

5. **Upload the data files and stored procedure files to ECS.**

   The example in this topic describes how to upload files to the following path:

   ```
   /tmp/dbName.sql
   ```

   ```
   /tmp/triggerProcedure.sql
   ```

6. **Log on to ECS and import the data files and stored procedure files to the target RDS for MySQL instance.**

   ```
   mysql -h intranet4example.mysql.rds.aliyuncs.com -u userName -p dbName
    < /tmp/dbName.sql
   ```

   ```
   mysql -h intranet4example.mysql.rds.aliyuncs.com -u userName -p dbName
    < /tmp/triggerProcedure.sql
   ```

   Parameter description:

   - **intranet4example.mysql.rds.aliyuncs.com: specifies the IP address that is used to connect to the RDS for MySQL instance. In this example, an intranet IP address is used.**
   - **userName: specifies the migration account of the RDS for MySQL database.**
   - **dbName: specifies the name of the database to be imported.**
   - **/tmp/dbName.sql: specifies the name of the data file to be imported.**
   - **/tmp/triggerProcedure.sql: specifies the name of the stored procedure file to be imported.**

9.15 Typical applications

9.15.1 Store multi-structure data

Apsara Stack Object Storage Service (OSS) is a cloud-based storage service that features large capacity, security, low costs, and high reliability. RDS and OSS can work together to implement a variety of data storage solutions.

Context

For example, when RDS and OSS are used in a forum, resources such as user images and forum posts can be stored in OSS, to reduce the storage pressure on RDS.

The following sample code enables the combined use of RDS and OSS.

Procedure

1. **Run the following command to initialize OssAPI:**

```
from oss.oss_api import * endpoint=" oss-cn-hangzhou.aliyuncs.com"
accessKeyId, accessKeySecret="your id","your secret" oss = OssAPI(
endpoint, accessKeyId, accessKeySecret)
```

2. **Run the following command to create a bucket:**

```
#Set the bucket ACL to Private: res = oss.create_bucket(bucket,"
private") print "%s\n%s" % (res.status, res.read())
```

3. **Run the following command to upload an object:**

```
res = oss.put_object_from_file(bucket, object, "test.txt") print "%s
\n%s" % (res.status, res.getheaders())
```

4. **Run the following command to obtain the corresponding object:**

```
res = oss.get_object_to_file(bucket, object, "/filepath/test.txt")
print "%s\n%s" % (res.status, res.getheaders())
```

In Elastic Compute Service (ECS) application code, the ID of each user is stored in RDS, and the avatar resources are stored in OSS. An example of Python code is as follows:

```
#! /usr/bin/env python from oss.oss_api import *
endpoint" oss-cn-hangzhou.aliyuncs.com"  accessKeyId, accessKeyS
ecret="your id", "your secret" oss = OssAPI(endpoint, accessKeyId,
accessKeySecret)
User_id = mysql_client.fetch_one (SQL) # Search for user_id in RDS.
#Obtain and download the user avatar to the corresponding path.
oss.get_object_to_file(bucket, object, your_path/user_id+'.png')
#Process the uploaded user avatar.
oss.put_object_from_file(bucket, object, your_path/user_id+'.png')
```

10 KVStore for Memcache

10.1 What is KVStore for Memcache?

KVStore for Memcache is a memory-based cache service for high-speed access to large amounts of small-sized data. KVStore for Memcache can reduce the load on back-end storage services and speed up website and application responses.

KVStore for Memcache supports data in the key-value structure. It can communicat e with Memcached-compatible clients.

KVStore for Memcache supports out-of-the-box deployment. It also relieves the load on databases from dynamic Web applications and improves website response speed by using the cache service.

Similar to user-created Memcached databases, KVStore for Memcache is also compatible with the Memcached protocol and user environments. The difference is that the data, hardware infrastructure, network security, and system maintenance services used by KVStore for Memcache are all deployed on the cloud.

10.2 Limits

Before using KVStore for Memcache, you need to understand the limits listed in the following table.

Item	Limit
Data type	KVStore for Memcache only supports the key-value data format. Complex data types such as array, map, and list are not supported .
Data reliability	The data of KVStore for Memcache is stored in the memory. Cached data cannot be guaranteed against data loss. KVStore for Memcache is not suitable to store data that requires high consistency.
Data amount	KVStore for Memcache supports a maximum of 1 KB in key size and 1 MB in value size for a single piece of cached data. KVStore for Memcache is not suitable to store data that exceeds these limits.

Item	Limit
Transaction support	KVStore for Memcache does not support transactions. Data that requires transactions must be written directly to the database.
Scenario	When data access traffic is evenly distributed and there are no obvious hot or cold spots, many access requests cannot hit the cached data in KVStore for Memcache. Therefore, KVStore for Memcache does not effectively function as a database cache. Before you select a database cache, you must consider the data access requirements of your business model.
Data deletion policy	Each key in KVStore for Memcache expires at a point in time designated by the user. After expiration, the key becomes inaccessible. The space occupied by the expired key is not recycled immediately after expiration, but is recycled at 02:00 every day.
Data expiration policy	Like open-source memcached, KVStore for Memcache adopts the Least Recently Used (LRU) algorithm to determine whether data expires. Expired data is not deleted and the space occupied by the expired data is not recycled immediately after expiration, but is recycled by a background program periodically.
Connection	The KVStore for Memcache server does not automatically close idle client connections.
Data expiration	We recommend that you control and manage the key expiration time.

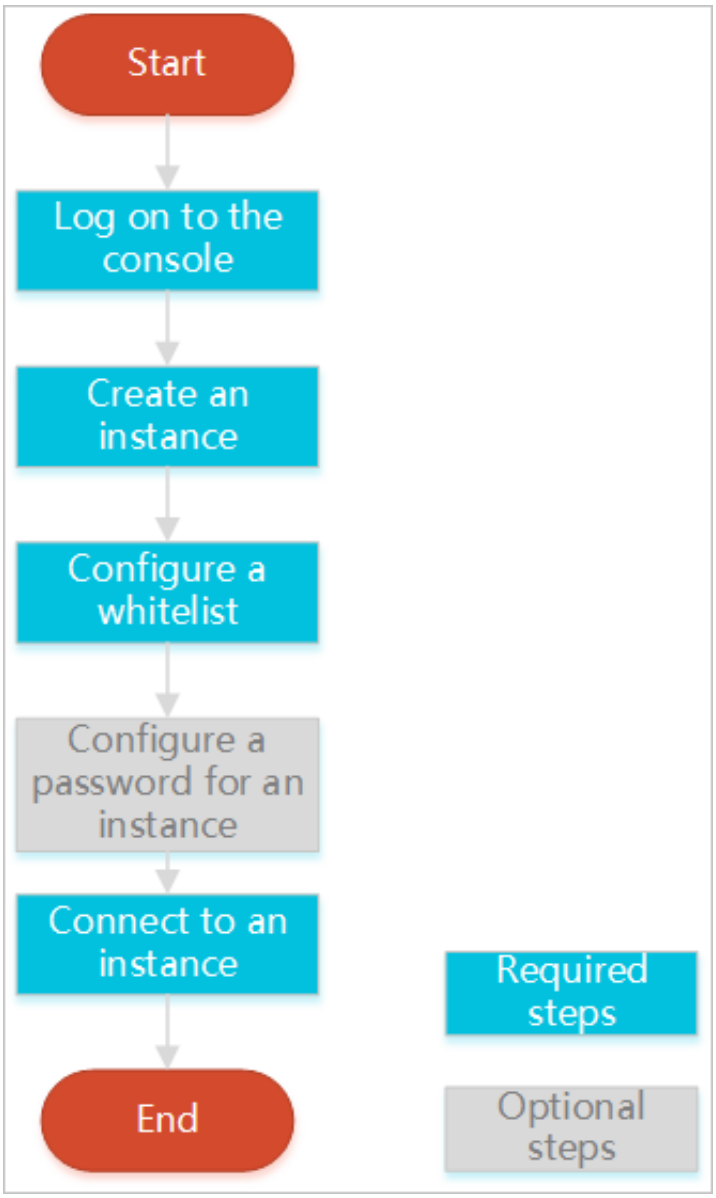
10.3 Quick start

10.3.1 Get started with KVStore for Memcache

This topic describes how to perform a series of operations such as creating an instance and log on to a database. This helps you understand how to operate a KVStore for Memcache instance.

For the detailed procedure, see *Figure 10-1: Flowchart of operating a KVStore for Memcache instance*.

Figure 10-1: Flowchart of operating a KVStore for Memcache instance



- *Log on to the KVStore for Memcache console*

  This topic describes how to log on to the KVStore for Memcache console.

- *Create an instance*

  KVStore for Memcache gives you the choice between a classic network or VPC. You can create KVStore for Memcache instances of different network types.

- *Configure a whitelist*

  To ensure a secure and stable database, you need to add IP addresses or IP address segments used for database access to the whitelist of a KVStore for Memcache instance before using the instance.

- *Reset the password*

  If you did not set a password for the instance when you created the instance, you must set a password on the Instance Information page.

- *Connect to an instance from a client*

  You can use a client that supports the memcached protocol to connect to an instance.

10.3.2 Log on to the KVStore for Memcache console

This topic describes how to log on to the KVStore for Memcache console.

Prerequisites

- **Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://`IP address or domain name of the Apsara Stack console`/manage.**
- **We recommend that you use the Chrome browser.**

Procedure

1. **Open your browser.**
2. **In the address bar, enter the access address of the Apsara Stack console in the format of http://`IP address or domain name of the Apsara Stack console`/manage, and then press Enter.**
3. **Enter the correct username and password.**
   - **The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.**
   - **You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following**

characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).

4. Click LOGIN to go to the Dashboard page.
5. In the top navigation bar, choose  > Database > KVStore.
6. Click the KVStore for Memcache tab.

10.3.3 Create an instance

KVStore for Memcache gives you the choice between a classic network or VPC. You can create KVStore for Memcache instances of different network types. This topic describes how to create an instance in the KVStore for Memcache console.

Prerequisites

- At least one ECS instance is required for activating KVStore for Memcache.
- To create a KVStore for Memcache instance of the VPC type, you must first create a VPC. Then, create the instance in the same region as the VPC.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. On the KVStore for Memcache tab, click Create Instance in the upper-right corner.
3. In the Create Memcache Instance page that appears, select a network type and specify the relevant settings.

Table 10-1: Parameter description

Category	Parameter	Description
Region	Region	The region in which you want to create a KVStore for Memcache instance. KVStore for Memcache allows only internal network access. Make sure the Memcache and ECS instances are in the same zone of the same region.

Category	Parameter	Description
	Zone	The zone in which you want to create a KVStore for Memcache instance. KVStore for Memcache allows only internal network access. Make sure the Memcache and ECS instances are in the same zone of the same region.
Basic Settings	Department	The department to which the KVStore for Memcache instance belongs.
	Project	The project to which the KVStore for Memcache instance belongs. **Notice:** After a project is selected, the KVStore for Memcache instance is accessible only to the members of the selected project. For more information, see View project members in *Apsara Stack Console User Guide* .
Instance Specification	Instance Specification	The instance specifications. The maximum connections and maximum internal network bandwidth vary depending on different instance specifications.

Category	Parameter	Description
Network	Network Type	The network type of the instance. On the Alibaba Cloud platform, a classic network and a VPC have the following differences: · Classic network: The cloud services in a classic network are not isolated. Unauthorized access can be blocked only by the security group or whitelist policy of the cloud services. · Virtual Private Cloud (VPC): A VPC helps you build an isolated network environment in Alibaba Cloud. You can customize the routing table, IP address range, and gateway in a VPC. In addition, you can combine your on-premises IDC with cloud resources in the Alibaba Cloud VPC through a leased line or VPN to migrate applications smoothly to the cloud. You must first create a VPC before you can set the network type to VPC. For more information, see Create a VPC and a VSwitch in *VPC User Guide* .
Password	Set Password	The password used to access the instance. You can select Now to set the password immediately or Later to set the password after creation. For more information, see *Reset the password*. The password must meet the following requirements: · It can be 8 to 30 characters in length and must contain uppercase letters, lowercase letters, and digits. · Special characters are not supported.

Category	Parameter	Description
Instance Name	Instance Name	The instance name. The name must be 2 to 128 characters in length. Spaces and special characters are not supported. Special characters include @/:="<>{[]}

4. Click Create.

   After creating the instance, wait until the instance status becomes Normal.

10.3.4 Configure a whitelist

To ensure a secure and stable database, you need to add IP addresses or IP address segments used for database access to the whitelist of a KVStore for Memcache instance before using the instance. This topic describes how to configure a whitelist.

Context

The whitelist improves access security for KVStore for Memcache. We recommend that you update the whitelist on a regular basis.

Notice:

- The ECS instance whose IP address is added to the whitelist must be in the same region as the KVStore for Memcache instance.
- To enable applications to access multiple KVStore for Memcache instances from the same ECS instance, bind the IP address of the ECS instance to multiple KVStore for Memcache instances.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, click the instance ID or click the  icon in the Actions column and choose View Details from the shortcut menu. The Instance Information page appears.
3. Click the Security Settings tab.

4. Click Create Whitelist Group or the modify icon next to the default whitelist group.
5. In the Change Whitelist dialog box that appears, configure parameters as prompted.

   Enter the IP addresses or IP address segments that are allowed to access the KVStore for Memcache instance. To allow all IP addresses to access the instance, set the whitelist to 0.0.0.0/0. To disable instance access from all IP addresses, set the whitelist to 127.0.0.1. We recommend that you delete the default IP address 127.0.0.1 from the whitelist. If you do not delete this IP address, other IP addresses or IP address segments that you add will not take effect.

   Notice:
   Separate multiple IP addresses or IP address segments with commas (no space before or after each comma), for example: 192.168.0.1,172.16.213.9. You can add up to 1,000 IP addresses to each whitelist.
6. Click OK.

10.3.5 Connect to an instance from a client

10.3.5.1 Overview

Any client that is compatible with the memcached protocol can access KVStore for Memcache. Each memcached client has its own characteristics. You can select any memcached client that supports Simple Authentication and Security Layer (SASL) and the memcached binary protocol based on your application characteristics.

The memcached clients described in the following topics can interact smoothly with KVStore for Memcache, and therefore are recommended for use.

Notice:

The third-party open-source clients described in the following topics are not provided by Alibaba Cloud and may contain defects. You must ensure the quality of the clients. Alibaba Cloud is not liable for any direct or indirect faults or losses caused by third-party clients.

10.3.5.2 Java: Spymemcache

You can use Java: Spymemcache to connect to a KVStore for Memcache instance.

Context

Download the client

- *Download address*
- *About the client*
- *Client versions*

Sample code

Procedure

1. **Prepare the Java development environment. Log on to an Alibaba Cloud ECS instance, and install the Java Development Kit (JDK) and a commonly used integrated development environment (IDE) such as Eclipse on the instance.**
   - **Java JDK** *Download address*
   - **Eclipse** (*Download address 1*, *Download address 2*)
2. **The first sample code is as follows. Copy the Java code to the Eclipse project.**

   Note:
   You must download a JAR package from a third party to call the KVStore for Memcache cache service. Otherwise, you cannot compile the code. After a JAR package is added, the code can be compiled.

   OcsSample1.java sample code (username and password required)

```
import java.io.IOException;
 import java.util.concurrent.ExecutionException;
 import net.spy.memcached.AddrUtil;
 import net.spy.memcached.ConnectionFactoryBuilder;
 import net.spy.memcached.ConnectionFactoryBuilder.Protocol;
 import net.spy.memcached.MemcachedClient;
 import net.spy.memcached.auth.AuthDescriptor;
 import net.spy.memcached.auth.PlainCallbackHandler;
 import net.spy.memcached.internal.OperationFuture;
 public class OcsSample1 {
      public static void main(String[] args) {
                final String host = "xxxxxxxx.m.yyyyyyyyyy.ocs.
aliyuncs.com";// The internal network address displayed in the
console.
                final String port =”11211”; // Default port: 11211.
You do not need to modify it.
                final String username = "xxxxxxxxx";// The access
account displayed in the console.
                final String password = "my_password";// The
password provided in the email.
```

```
                MemcachedClient cache = null;
                try {
                        AuthDescriptor ad = new AuthDescriptor(new
String[]{"PLAIN"}, new PlainCallbackHandler(username, password));
                        cache = new MemcachedClient(
                                                new ConnectionFactoryBui
lder().setProtocol(Protocol.BINARY)
                                .setAuthDescriptor(ad)
                                .build(),
                                AddrUtil.getAddresses(host + ":" +
port));
                        System.out.println("OCS Sample Code");
                        // Save a value with the "ocs" key to
KVStore for Memcache to facilitate data verification and reading.
                        String key = "ocs";
                        String value = "Open Cache Service,  from
www.Aliyun.com";
                        int expireTime = 1000; // Expiration time
, unit: seconds. The countdown starts once data is written. After
expireTime is exceeded, the data expires and cannot be read.
                        OperationFuture<Boolean> future = cache.set
(key, expireTime, value);
                        future.get();  // The spymemcached set()
method is asynchronous. The future.get() operation starts after the
 cache.set() operation is completed. You can also choose to execute
both operations at the same time.
                        // Save several values to KVStore for
Memcache and you can view the statistics in the KVStore for Memcache
 console.
                        for(int i=0;i<100;i++){
                               key="key-"+i;
                               value="value-"+i;
                               // Perform the Set operation and
save the value to the cache.
                               expireTime = 1000; // Expiration
time, unit: seconds.
                               future = cache.set(key, expireTime,
value);
                               future.get();  // Make sure that the
 previous cache.set() operation is completed.
                        }
                        System.out.println(”Set operation completed
!") ;
                           // Perform the Get operation and read
the value with the "ocs" key from the cache.
                        System.out.println("Get operation: "+cache.
get(key));
                               } catch (IOException e) {
                                      e.printStackTrace();
                               } catch (InterruptedException e) {
                                      e.printStackTrace();
                               } catch (ExecutionException e) {
                                      e.printStackTrace();
                               }
                               if (cache ! = null) {
                                      cache.shutdown();
                               }
      }//eof
 }
```

OcsSample2.java sample code (username and password not required)

```
import java.io.IOException;
```

```
 import java.util.concurrent.ExecutionException;
 import net.spy.memcached.AddrUtil;
 import net.spy.memcached.BinaryConnectionFactory;
 import net.spy.memcached.MemcachedClient;
 import net.spy.memcached.internal.OperationFuture;
 public class OcsSample2 {
 public static void main(String[] args) {
     final String host = "xxxxxxxx.m.yyyyyyyyyy.ocs.aliyuncs.com
"; // The internal network address displayed in the console
     final String port = "11211"; // Default port: 11211. You do not
 need to modify it.
     MemcachedClient cache = null;
     try {
         cache = new MemcachedClient(new BinaryConnectionFactory(),
AddrUtil.getAddresses(host + ":" + port));
         System.out.println("OCS Sample Code");
         // Save a value with the "ocs" key to KVStore for Memcache
to facilitate data verification and reading.
         String key = "ocs";
         String value = "Open Cache Service,  from www.Aliyun.com";
         int expireTime = 1000; // Expiration time, unit: seconds
. The countdown starts once data is written. After expireTime is
exceeded, the data expires and cannot be read.
         OperationFuture<Boolean> future = cache.set(key, expireTime
, value);
         future.get();
         // Save several values to KVStore for Memcache and you can
view the statistics in the KVStore for Memcache console.
         for (int i = 0; i < 100; i++) {
             key = "key-" + i;
             value = "value-" + i;
             // Perform the Set operation and save the value to the
cache.
             expireTime = 1000; // Expiration time, unit: seconds.
             future = cache.set(key, expireTime, value);
             future.get();
         }
         System.out.println(”Set operation completed!") ;
         // Perform the Get operation and read the value with the "
ocs" key from the cache.
         System.out.println("Get operation: " + cache.get(key));
     } catch (IOException e) {
         e.printStackTrace();
     } catch (InterruptedException e) {
         e.printStackTrace();
     } catch (ExecutionException e) {
         e.printStackTrace();
     }
     if (cache ! = null) {
         cache.shutdown();
     }
 }//eof
 }
```

3. **Open OcsSample1.java in Eclipse. Modify the instance ID and internal network address in OcsSample1.java based on your instance information.**

4. **After the information is modified, you can run your program. Run the main function. The following result is displayed in the console window within Eclipse (ignore the red INFO debugging information that may be displayed).**

```
OCS Sample Code
 Set operation completed!
 Get operation: Open Cache Service,  from www.Aliyun.com
```

10.3.5.3 PHP Memcached

You can use PHP Memcached to connect to a KVStore for Memcache instance.

Context

Download a client

- *Download addresses*
- *About the client*
- *Client versions*

System requirements and environment configuration

> **Notice:**
> **If you already have a PHP Memcache environment, pay attention to the tips in the tutorial. Otherwise, your production environment may be overwritten and services may become unavailable. We recommend that you back up your data before upgrading or compiling the environment.**

PHP Memcached for Windows

If the environment cannot be established using the standard PHP Memcached extensions, you can splice packets manually to access KVStore for Memcache. For connection methods, see the following link. The sample code is simple. In comparison with PHP Memcached, PHP Memcached for Windows only supports mainstream interfaces, so you need to perform additional operations to use it with other specific interfaces. For installation and usage methods, click *Here*.

PHP Memcached for CentOS or Aliyun Linux 6

> **Notice:**
> **Memcached 2.2.0 extensions must use libmemcached 1.0.x. Libraries earlier than 1.0 cannot be compiled. GCC 4.2 or later must be used to compile libmemcached.**

1. **Check whether components such as GCC-C++ are installed (run the `gcc -v` command to check whether GCC 4.2 or later is used). If not, run the `yum install gcc+ gcc-c++` command to install them.**
2. **Run the `rpm -qa | grep php` command to check whether the PHP environment is ready in the system. If not, run the `yum install php-devel php-common php-cli` command to install PHP with source code compiling.**

   PHP 5.3 or later is recommended. The PHP 5.2 source code contains the zend_parse_parameters_none function, which may cause errors. If you need to compile the source code, follow the official PHP compiling and upgrading methods.
3. **Check whether SASL-related dependencies are installed. If not, run the `yum install cyrus-sasl-plain cyrus-sasl cyrus-sasl-devel cyrus-sasl-lib` command to install SASL-related environments.**
4. **Check whether the libmemcached source code package is installed. If not, run the following commands to install it. libmemcached 1.0.18 is recommended.**

```
 wget https://launchpad.net/libmemcached/1.0/1.0.18/+download/
libmemcached-1.0.18.tar.gz
 tar zxvf libmemcached-1.0.18.tar.gz
 cd libmemcached-1.0.18
 ./configure --prefix=/usr/local/libmemcached --enable-sasl
 make
 make install
 cd ..
```

5. **Run the `yum install zlib-devel` command to install the Memcached source code package (Memcached 2.2.0 recommended).**

   Notice:

   - **Before installing Memcached, check whether there are any zlib-devel packages to be executed.**
   - **You must first check whether the Memcached client package (including the source code package) is installed. If yes, recompile it to add the -enable-memcached-sasl extension.**

```
wget http://pecl.php.net/get/memcached-2.2.0.tgz
tar zxvf memcached-2.2.0.tgz
cd memcached-2.2.0
phpize (If there are two sets of PHP environments in the system,
you must enter the absolute path /usr/bin/phpize. The path is the
 PHP environment path for using KVStore for Memcache.)
```

```
./configure --with-libmemcached-dir=/usr/local/libmemcached --
enable-memcached-sasl (Pay special attention to this parameter.)
make
make install
```

6. **Modify the php.ini file. Run the locate command to find this file. If there are two sets of PHP environments in the system, you must locate the PHP environment path for using KVStore for Memcache and modify it accordingly. Add `extension=memcached.so memcached.use_sasl = 1`.**
7. **Test whether the production environment is successfully deployed by using the test code provided at the end of the page. Replace the address, port number, username, and password in the test code with actual ones.**

PHP Memcached for CentOS or Aliyun Linux 5 (64-bit)

1. **Check whether components such as GCC-C++ are installed. If not, run the `yum install gcc+ gcc-c++` command to install them.**
2. **Run the `rpm -qa | grep php` command to check whether the PHP environment is ready in the system. If not, run the `yum install php53 php53-devel` command to install PHP with source code compiling. If the PHP environment has been prepared, skip this step. PHP 5.3 or later is recommended.**

   The PHP 5.2 source code contains the zend_parse_parameters_none function, which may cause errors.
3. **Run the `yum install cyrus-sasl-plain cyrus-sasl cyrus-sasl-devel cyrus-sasl-lib` command to install SASL-related environments.**
4. **Check whether libmemcached (including the source code package) is installed. libmemcached 1.0.2 is recommended. If not, run the following commands to install libmemcached:**

```
 wget http://launchpad.net/libmemcached/1.0/1.0.2/+download/
libmemcached-1.0.2.tar.gz
 tar -zxvf libmemcached-1.0.2.tar.gz
 cd libmemcached-1.0.2
 ./configure --prefix=/usr/local/libmemcached --enable-sasl
 make
 make install
 cd ..
```

5. **Run the `yum install zlib-devel` command to install the Memcached source code package (Memcached 2.0 recommended).**

Note:

- **Before installing Memcached, check whether there are any zlib-devel packages to be executed.**
- **You must first check whether the Memcached client package (including the source code package) is installed. If yes, recompile it to add the -enable-memcached-sasl extension.**

```
wget http://pecl.php.net/get/memcached-2.0.0.tgz tar -zxvf
memcached-2.0.0.tgz
cd memcached-2.0.0
phpize (If there are two sets of PHP environments in the system
, you must enter the absolute path /usr/bin/phpize. The path is
the PHP environment path for using KVStore for Memcache. Run the
phpize command in the Memcached source code directory.)
./configure --with-libmemcached-dir=/usr/local/libmemcached --
enable-memcached-sasl (Pay special attention to this parameter.)
make
make install
```

6. **Modify the php.ini file. Run the locate command to find this file. Files installed with the yum command are in /etc/php.ini. If there are two sets of PHP environments in the system, you must locate the PHP environment path for using KVStore for Memcache and modify it accordingly. Add `extension=memcached.so` `memcached.use_sasl = 1`.**
7. **Run the `php -m |grep memcached` command. If the displayed result contains "memcache", KVStore for Memcache is supported.**
8. **Test whether the production environment is successfully deployed by using the test code provided at the end of the page. Replace the address, port number, username, and password in the test code with actual ones.**

PHP Memcached for Ubuntu Debian

1. **Change the Ubuntu source.**

   Solution 1: Run the `vim /etc/apt/source.list` command and add the following content at the beginning of the file:

```
deb http://mirrors.aliyun.com/ubuntu/ precise main restricted
universe multiverse
 deb http://mirrors.aliyun.com/ubuntu/ precise-security main
restricted universe multiverse
 deb http://mirrors.aliyun.com/ubuntu/ precise-updates main
restricted universe multiverse
 deb http://mirrors.aliyun.com/ubuntu/ precise-proposed main
restricted universe multiverse
 deb http://mirrors.aliyun.com/ubuntu/ precise-backports main
restricted universe multiverse
 deb-src http://mirrors.aliyun.com/ubuntu/ precise main restricted
universe multiverse
```

```
 deb-src http://mirrors.aliyun.com/ubuntu/ precise-security main
restricted universe multiverse
 deb-src http://mirrors.aliyun.com/ubuntu/ precise-updates main
restricted universe multiverse
 deb-src http://mirrors.aliyun.com/ubuntu/ precise-proposed main
restricted universe multiverse
 deb-src http://mirrors.aliyun.com/ubuntu/ precise-backports main
restricted universe multiverse
 apt-get update // Update the list.
```

Solution 2: Run the `wget http://oss.aliyuncs.com/aliyunecs/update_source.zip` command to download the update_source package. Extract the package. Run the `chmod 777 file name` command to grant the file execution permission. Run the script to automatically change the source.

2. **Run the `ape-get` command to configure GCC-C++.**

   You must first run the `dpkg -s installation package name` command (such as `dpkg -s gcc`) to check whether components such as GCC-C++ are installed. If not, run the `apt-get build-dep gcc apt-get install build-essential` command.

3. **Install php5 and php5-dev.**

   You must first run the `dpkg -s installation package name` command (such as `dpkg -s php`) to check whether components such as PHP are installed. If not, run `apt-get install php5 php5-dev` command. php5-cli and php5-common are automatically installed at the same time.

4. **Install and configure SASL-related dependencies.**

   You must first run the `dpkg -s installation package name` command (such as `dpkg -s libsasl2`) to check whether components such as cloog-ppl are installed. If not, run the following commands to install them:

   ```
   apt-get install libsasl2-dev cloog-ppl
   cd /usr/local/src
   ```

5. **Run the following commands to install libmemcached of the specified version:**

   Notice:
   Before running the commands, check whether the specified package (including the source code package) is installed. If yes, skip this step.

   ```
    wget https://launchpad.net/libmemcached/1.0/1.0.18/+download/
   libmemcached-1.0.18.tar.gz
    tar -zxvf libmemcached-1.0.18.tar.gz
    cd libmemcached-1.0.18
   ```

```
 ./configure --prefix=/usr/local/libmemcached
 make
 make install
 cd ..
```

6. **Run the following commands to install Memcached of the specified version.**

   Notice:
   You must first check whether the Memcached client package (including the source package) is installed. If the Memcached client package has been installed, recompile it to add the -enable-memcached-sasl extension.

```
 wget
 http://pecl.php.net/get/memcached-2.2.0.tgz
tar zxvf memcached-2.2.0.tgz
 cd memcached-2.2.0 phpize5
 ./configure --with-libmemcached-dir=/usr/local/libmemcached --
enable-memcached-sasl
 make
 make install
```

7. **Configure PHP to support Memcached and then test the configuration.**

```
 echo "extension=memcached.so" >>/etc/php5/conf.d/pdo.ini
echo "memcached.use_sasl = 1" >>/etc/php5/conf.d/pdo.ini
php -m |grep mem memcached
```

   If this component is displayed, the installation and configuration are complete.

Sample code

Example 1: Basic commands to connect to KVStore for Memcache and set and get commands

```
<? php
$connect = new Memcached; // Declare a new memcached connection.
$connect->setOption(Memcached::OPT_COMPRESSION, false); // Disable the
 compression function.
$connect->setOption(Memcached::OPT_BINARY_PROTOCOL, true); // Use the
binary protocol.
$connect->setOption(Memcached::OPT_TCP_NODELAY, true); // Note: PHP
Memcached has a bug that causes a fixed latency of 40 ms when there is
 no Get value. Set this parameter to true to prevent this bug.
$connect->addServer('aaaaaaaaaa.m.yyyyyyyyyy.ocs.aliyuncs.com', 11211
); // Add the address and port number of the KVStore for Memcache
instance.
$connect->setSaslAuthData('aaaaaaaaaa', 'password'); // Set the
KVStore for Memcache account and password for authentication. Skip
this step if the password-free function is enabled.
$connect->set("hello", "world");
echo 'hello: ',$connect->get("hello");
$connect->quit();
```

```
? >
```

Example 2: Cache an array in KVStore for Memcache

```
<? php
$connect= new Memcached; // Declare a new Memcached connection.
$connect->setOption(Memcached::OPT_COMPRESSION, false); // Disable the
 compression function.
$connect->setOption(Memcached::OPT_BINARY_PROTOCOL, true);// Use the
binary protocol.
$connect->setOption(Memcached::OPT_TCP_NODELAY, true); //Note: PHP
Memcached has a bug that causes a fixed latency of 40 ms when there is
 no Get value. Set this parameter to true to prevent this bug.
$connect->addServer('xxxxxxxx.m.yyyyyyyy.ocs.aliyuncs.com', 11211);//
Add the address and port number of the KVStore for Memcache instance.
$connect->setSaslAuthData('xxxxxxxx', 'bbbbbbbb');// Set the KVStore
for Memcache account and password for authentication. Skip this step
if the password-free function is enabled.
$user = array(
    "name" => "ocs",
    "age" => 1,
    "sex" => "male"
); // Declare an array.
$expire = 60; // Set an expiration time.
test($connect->set('your_name',$user,$expire), true, 'Set cache failed
');
if($connect->get('your_name')){
$result =$connect->get('your_name');
}else{
echo "Return code:", $connect->getResultCode();
echo "Return Message:", $connect->getResultMessage (); // If an error
is returned, parse the return code.
$result=" ";
}
print_r($result);
$connect->quit();
function test($val, $expect, $msg)
{
    if($val! = $expect) throw new Exception($msg);
}
? >
```

Example 3: Use KVStore for Memcache combined with a MySQL database

```
<? php
$connect = new Memcached; // Declare a new Memcached connection.
$connect->setOption(Memcached::OPT_COMPRESSION, false);// Disable the
compression function.
$connect->setOption(Memcached::OPT_BINARY_PROTOCOL, true);// Use the
binary protocol.
$connect->setOption(Memcached::OPT_TCP_NODELAY, true); //Note: PHP
Memcached has a bug that causes a fixed latency of 40 ms when there is
 no Get value. Set this parameter to true to prevent this bug.
$connect->addServer('xxxxxx.m.yyyyyyyy.ocs.aliyuncs.com', 11211);//
Add the instance address and port number.
$connect->setSaslAuthData('xxxxxx', 'my_passwd');// Set the KVStore
for Memcache account and password for authentication. Skip this step
if the password-free function is enabled.
$user = array(
   "name" => "ocs",
   "age"  => 1,
   "sex"  => "male"
```

```
); // Define an array.
if($connect->get('your_name'))
{
  $result =$connect->get('your_name');
  print_r($result);
  echo "Found in OCS, get data from OCS"; //If the value is obtained,
the value source is displayed as KVStore for Memcache.
  exit;
}
else
{
  echo "Return code:", $connect->getResultCode();
  echo "Return Message:", $connect->getResultMessage ();// Send the
return code.
  $db_host='zzzzzz.mysql.rds.aliyuncs.com'; // Database address.
  $db_name='my_db';          // Database name.
  $db_username='db_user';    // Database username.
  $db_password='db_passwd'; // Database password.
  $connection=mysql_connect($db_host,$db_username,$db_password);
  if (! mysql_select_db($db_name, $connection))
  {
    echo 'Could not select database'; // An error is returned if the
database connection fails.
    exit;
  }
  $sql = "SELECT name,age,sex FROM test1 WHERE name = 'ocs'";
  $result = mysql_query($sql, $connection);
  while ($row = mysql_fetch_assoc($result))
  {
    $user = array(
      "name" => $row["name"],
      "age"  => $row["age"],
      "sex"  => $row["sex"],
    );
    $expire = 5; // Set the expiration time of data in the cache.
    test($connect->set('your_name',$user,$expire), true, 'Set cache
failed'); // Write the value to the KVStore for Memcache cache.
  }
  mysql_free_result($result);
  mysql_close($connection);
}
print_r($connect->get('your_name')); // Return the value obtained.
echo "Not Found in OCS,get data from MySQL"; // Confirm the value
obtained from the database.
$connect->quit();
function test($val, $expect, $msg)
{
  if($val! = $expect) throw new Exception($msg);
}
? >
```

10.3.5.4 Python

You can use Python to connect to a KVStore for Memcache instance.

Download the client

- *Download address*
- *About the client*
- *Client versions*

Environment configuration

The bmemcached dependency that supports SASL must be installed. To download bmemcached, click *here*.

Sample code

```
#! /usr/bin/env python
import bmemcached
client = bmemcached.Client(('ip:port'), 'user', 'passwd')
print client.set('key', 'value11111111111')
print client.get('key')
```

10.3.5.5 C#/.NET: EnyimMemcached

You can use C#/.NET: EnyimMemcached to connect to a KVStore for Memcache instance.

Download the client

- *Download address*
- *About the client*
- *Client versions*

Sample code

```
using System.Net;
using Enyim.Caching;
using Enyim.Caching.Configuration;
using Enyim.Caching.Memcached;
namespace OCS.Memcached
{
    public sealed class MemCached
    {
        private static MemcachedClient MemClient;
        static readonly object padlock = new object();
        // Thread-safe single instance mode.
        public static MemcachedClient getInstance()
        {
            if (MemClient == null)
            {
                lock (padlock)
                {
                    if (MemClient == null)
                    {
                        MemClientInit();
                    }
                }
            }
            return MemClient;
        }
        static void MemClientInit()
        {
            // Initialize the cache.
            MemcachedClientConfiguration memConfig = new MemcachedC
lientConfiguration();
```

```
            IPAddress newaddress =
   IPAddress.Parse(Dns.GetHostEntry
 ("your_ocs_host").AddressList[0].ToString()); // Replace your_ocs_host
with the Memcache internal network address.
            IPEndPoint ipEndPoint = new IPEndPoint(newaddress, 11211);
              // Configuration file - IP address.
            memConfig.Servers.Add(ipEndPoint);
            // Configuration file - protocol.
           memConfig.Protocol = MemcachedProtocol.Binary;
            // Configuration file - permission.
            memConfig.Authentication.Type = typeof(PlainTextA
uthenticator);
            memConfig.Authentication.Parameters["zone"] = "";
            memConfig.Authentication.Parameters["userName"] = "
username";
            memConfig.Authentication.Parameters["password"] = "
password";
      // Complete the following settings based on the maximum
connections of the instance.
            memConfig.SocketPool.MinPoolSize = 5;
            memConfig.SocketPool.MaxPoolSize = 200;
            MemClient=new MemcachedClient(memConfig);
                  }
          }
}
```

Dependency

Code:

```
MemcachedClient MemClient = MemCached.getInstance();
```

10.3.5.6 C++

You can use a C++ client to connect to a KVStore for Memcache instance.

Download the client

- *Download address*
- *About the client*
- *Client versions*

Environment configuration

1. **Download, compile, and install the C++ client.**

   https://launchpad.net/libmemcached/1.0/1.0.18/+download/libmemcached-1.0.18.tar.gz

2. **Run the following commands:**

   ```
   tar -xvf libmemcached-1.0.18.tar.gz

   cd libmemcached-1.0.18

   ./configure

   sudo make install
   ```

C++ sample code

1. **Download** *ocs_test.tar.gz*.
2. **Run the following commands:**

   ```
   tar -xvf ocs_test.tar.gz

   cd ocs_test

   vim ocs_test_sample1.cpp
   ```

3. **Set TARGET_HOST to the internal network address of the KVStore for Memcache instance, USERNAME to the username of your instance, and PASSWORD to the password you set.**
4. **Run the `build.sh` command to generate ocs_test. Run the `./ocs_test` command. A key is written to the KVStore for Memcache instance. Get the key from the Memcache instance and delete the key from the instance.**

   The code of ocs_test_sample1.cpp is as follows:

   ```
   #include <iostream>
    #include <string>
    #include <libmemcached/memcached.h>
    using namespace std;
    #define TARGET_HOST  ""
    #define USERNAME ""
    #define PASSWORD ""
    int main(int argc, char *argv[])
    {
        memcached_st *memc = NULL;
        memcached_return rc;
        memcached_server_st *server;
        memc = memcached_create(NULL);
        server = memcached_server_list_append(NULL, TARGET_HOST, 11211
   ,&rc);
        /* SASL */
        sasl_client_init(NULL);
        rc = memcached_set_sasl_auth_data(memc, USERNAME, PASSWORD);
        if(rc ! = MEMCACHED_SUCCESS) {
            cout<<"Set SASL err:"<< endl;
        }
        rc = memcached_behavior_set(memc,MEMCACHED_BEHAVIOR_BINARY_PROT
   OCOL,1);
   ```

```
    if(rc ! = MEMCACHED_SUCCESS) {
        cout<<"Binary Set err:"<<endl;
    }
    /* SASL */
    rc = memcached_server_push(memc,server);
    if(rc ! = MEMCACHED_SUCCESS) {
      cout <<"Connect Mem err:"<< rc << endl;
    }
    memcached_server_list_free(server);
    string key = "TestKey";
    string value = "TestValue";
    size_t value_length = value.length();
    size_t key_length = key.length();
    int expiration = 0;
    uint32_t flags = 0;
    //Save data
    rc = memcached_set(memc,key.c_str(),key.length(),value.c_str(),
value.length(),expiration,flags);
    if (rc ! = MEMCACHED_SUCCESS){
      cout <<"Save data failed: " << rc << endl;
      return -1;
    }
    cout <<"Save data succeed, key: " << key << " value: " << value
 << endl;
    cout << "Start get key:" << key << endl;
    char* result = memcached_get(memc,key.c_str(),key_length,&
value_length,&flags,&rc);
    cout << "Get value:" << result << endl;
    //Delete data
    cout << "Start delete key:" << key << endl;
    rc = memcached_delete(memc,key.c_str(),key_length,expiration);
    if (rc ! = MEMCACHED_SUCCESS) {
      cout << "Delete key failed: " << rc << endl;
    }
    cout << "Delete key succeed: " << rc << endl;
    //free
    memcached_free(memc);
    return 0;
 }
```

The following example shows the use of KVStore for Memcache in a different C++ client, where KVStore for Memcache and MySQL are combined. You can follow the steps in the preceding example to compile and install the C++ client.

1. **Create a sample database and a table in the MySQL database.**

```
mysql -h host -P port -u USER -p PASSWORD

create database testdb;

create table user_info (user_id int, user_name char(32) not null,
password char(32) not null, is_online int, primary key(user_id) );
```

2. **Download** *ocs_test_2.tar.gz* **and run the following commands:**

```
tar -xvf ocs_test_2.tar.gz

cd ocs_test

vim ocs_test_sample2.cpp
```

Note:

Set OCS_TARGET_HOST to the internal network address of the KVStore for Memcache instance, OCS_USERNAME to the KVStore for Memcache instance name, OCS_PASSWORD to the password that you set, MYSQL_HOST to the MySQL database address, MYSQL_USERNAME to the database username, and MYSQL_PASSWORD to the database password.

3. **Run the** `build.sh` **command to generate ocs_test and run the** `./ocs_test` **command.**

The code of ocs_test_sample2.cpp is as follows:

```
#include <iostream>
 #include <string>
 #include <sstream>
 #include <libmemcached/memcached.h>
 #include <mysql/mysql.h>
 using namespace std;
 #define OCS_TARGET_HOST  "xxxxxxxxxx.m.yyyyyyyyy.ocs.aliyuncs.com"
 #define OCS_USERNAME "your_user_name"
 #define OCS_PASSWORD "your_password"
 #define MYSQL_HOST         "zzzzzzzzzz.mysql.rds.aliyuncs.com"
 #define MYSQL_USERNAME     "db_user"
 #define MYSQL_PASSWORD     "db_paswd"
 #define MYSQL_DBNAME       "testdb"
 #define TEST_USER_ID       "100"
 MYSQL *mysql = NULL;
 memcached_st *memc = NULL;
 memcached_return rc;
 int InitMysql()
 {
   mysql = mysql_init(0);
   if (mysql_real_connect(mysql, MYSQL_HOST, MYSQL_USERNAME,
MYSQL_PASSWORD, MYSQL_DBNAME, MYSQL_PORT, NULL, CLIENT_FOUND_ROWS)
 == NULL )
   {
     cout << "connect mysql failure!" << endl;
     return EXIT_FAILURE;
   }
   cout << "connect mysql success!" << endl;
   return 0;
 }
 bool InitMemcached()
 {
   memcached_server_st *server;
   memc = memcached_create(NULL);
```

```
   server = memcached_server_list_append(NULL, OCS_TARGET_HOST,
11211,&rc);
   /* SASL */
   sasl_client_init(NULL);
   rc = memcached_set_sasl_auth_data(memc, OCS_USERNAME, OCS_PASSWO
RD);
   if (rc ! = MEMCACHED_SUCCESS)
   {
     cout<<"Set SASL err:"<< endl;
     return false;
   }
   rc = memcached_behavior_set(memc,MEMCACHED_BEHAVIOR_BINARY_PROT
OCOL,1);
   if (rc ! = MEMCACHED_SUCCESS)
   {
     cout<<"Binary Set err:"<<endl;
     return false;
   }
   /* SASL */
   rc = memcached_server_push(memc,server);
   if (rc ! = MEMCACHED_SUCCESS)
   {
     cout <<"Connect Mem err:"<< rc << endl;
     return false;
   }
   memcached_server_list_free(server);
   return true;
 }
 struct UserInfo
 {
   int user_id;
   char user_name[32];
   char password[32];
   int is_online;
 };
 bool SaveToCache(string &key, string &value, int expiration)
 {
   size_t value_length = value.length();
   size_t key_length = key.length();
   uint32_t flags = 0;
   //Save data
   rc = memcached_set( memc,key.c_str(), key.length(), value.c_str
(), value.length(), expiration, flags);
   if (rc ! = MEMCACHED_SUCCESS){
       cout <<"Save data to cache failed: " << rc << endl;
       return false;
   }
   cout <<"Save data to cache succeed, key: " << key << " value: "
 << value << endl;
   return true;
 }
 UserInfo *GetUserInfo(int user_id)
 {
   UserInfo *user_info = NULL;
   //get from cache
   string key;
   stringstream out;
   out << user_id;
   key = out.str();
   cout << "Start get key:" << key << endl;
   size_t value_length;
   uint32_t flags;
   char* result = memcached_get(memc, key.c_str(), key.size(), &
value_length, &flags, &rc);
```

```
   if (rc ! = MEMCACHED_SUCCESS)
   {
     cout << "Get Cache Failed, start get from mysql."<< endl;
     int status;
     char select_sql[1024];
     memset(select_sql, 0x0, sizeof(select_sql));
     sprintf(select_sql, "select * from user_info where user_id = %d
", user_id);
     status = mysql_query(mysql, select_sql);
     if (status ! =0 )
     {
       cout << "query from mysql failure!" << endl;
       return NULL;
     }
     cout << "the status is :" << status << endl;
     MYSQL_RES *mysql_result = mysql_store_result(mysql);
     user_info = new UserInfo;
     MYSQL_ROW row;
     while (row = mysql_fetch_row(mysql_result))
     {
       user_info->user_id = atoi(row[0]);
       strncpy(user_info->user_name, row[1], strlen(row[1]));
       strncpy(user_info->password, row[2], strlen(row[2]));
       user_info->is_online = atoi(row[3]);
     }
     mysql_free_result(mysql_result);
     return user_info;
   }
   cout << "Get from cache succeed" << endl;
   user_info = new UserInfo;
   memcpy(user_info, result, value_length);
   return user_info;
 }
 bool DeleteCache(string &key, int expiration)
 {
   rc = memcached_delete(memc, key.c_str(), key.length(), expiration
);
   if (rc ! = MEMCACHED_SUCCESS) {
     cout << "Delete key failed: " << rc << endl;
     return false;
   }
   cout << "Delete key succeed: " << rc << endl;
   return true;
 }
 void PrintUserInfo(UserInfo *user_info)
 {
   cout << "user_id: " << user_info->user_id << " " << " name: " <<
user_info->user_name << endl;
 }
 bool SaveMysql(UserInfo *user_info)
 {
   char insert_sql[1024];
   memset(insert_sql, 0x0, sizeof(insert_sql));
   sprintf(insert_sql, "insert into user_info(user_id, user_name,
password, is_online) values(%d, '%s', '%s', %d)", user_info->user_id
, user_info->user_name, user_info->password, user_info->is_online);
   int status = mysql_query(mysql, insert_sql);
   if (status ! = 0)
   {
     cout << "insert failed" << endl;
     return false;
   }
   cout << "insert user_info" << endl;
   //insert mysql
```

```
  return true;
}
int main(int argc, char *argv[])
{
  if (InitMysql() ! = 0)
  {
    return -1;
  }
  if (! InitMemcached())
  {
    return -1;
  }
  //generate user_info
  UserInfo user_info;
  user_info.user_id = atoi(TEST_USER_ID);
  strcpy(user_info.user_name, "James");
  strcpy(user_info.password, "12345678");
  user_info.is_online = 1;
  //save to mysql
  if (! SaveMysql(&user_info))
  {
    //return -1;
  }
  string user_str;
  user_str.assign((char*)&user_info, sizeof(UserInfo));
  //save to memcached
  string key_str = TEST_USER_ID;
  SaveToCache(key_str, user_str, 10);
  //start get, exist in memcahced
  UserInfo *get_user_info = GetUserInfo(user_info.user_id);
  PrintUserInfo(get_user_info);
  //wait 10 seconds
  sleep(2);
  //delete memcached or expired
  DeleteCache(key_str, 0);
  //start get, exist in mysql
  delete get_user_info;
  get_user_info = GetUserInfo(user_info.user_id);
  PrintUserInfo(get_user_info);
  delete get_user_info;
  //free
  memcached_free(memc);
  mysql_close(mysql);
  return 0;
}
```

10.4 Manage instances

10.4.1 Create an instance

KVStore for Memcache gives you the choice between a classic network or VPC. You can create KVStore for Memcache instances of different network types. This topic describes how to create an instance in the KVStore for Memcache console.

Prerequisites

- **At least one ECS instance is required for activating KVStore for Memcache.**

- To create a KVStore for Memcache instance of the VPC type, you must first create a VPC. Then, create the instance in the same region as the VPC.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. On the KVStore for Memcache tab, click Create Instance in the upper-right corner.
3. In the Create Memcache Instance page that appears, select a network type and specify the relevant settings.

   Table 10-2: Parameter description

Category	Parameter	Description
Region	Region	The region in which you want to create a KVStore for Memcache instance. KVStore for Memcache allows only internal network access. Make sure the Memcache and ECS instances are in the same zone of the same region.
	Zone	The zone in which you want to create a KVStore for Memcache instance. KVStore for Memcache allows only internal network access. Make sure the Memcache and ECS instances are in the same zone of the same region.
Basic Settings	Department	The department to which the KVStore for Memcache instance belongs.
	Project	The project to which the KVStore for Memcache instance belongs. **Notice:** After a project is selected, the KVStore for Memcache instance is accessible only to the members of the selected project. For more information, see View project members in *Apsara Stack Console User Guide* .

Category	Parameter	Description
Instance Specification	Instance Specification	The instance specifications. The maximum connections and maximum internal network bandwidth vary depending on different instance specifications.
Network	Network Type	The network type of the instance. On the Alibaba Cloud platform, a classic network and a VPC have the following differences: · Classic network: The cloud services in a classic network are not isolated. Unauthorized access can be blocked only by the security group or whitelist policy of the cloud services. · Virtual Private Cloud (VPC): A VPC helps you build an isolated network environment in Alibaba Cloud. You can customize the routing table, IP address range, and gateway in a VPC. In addition, you can combine your on-premises IDC with cloud resources in the Alibaba Cloud VPC through a leased line or VPN to migrate applications smoothly to the cloud. You must first create a VPC before you can set the network type to VPC. For more information, see Create a VPC and a VSwitch in *VPC User Guide* .

Category	Parameter	Description
Password	Set Password	The password used to access the instance. You can select Now to set the password immediately or Later to set the password after creation. For more information, see *Reset the password*. The password must meet the following requirements: · It can be 8 to 30 characters in length and must contain uppercase letters, lowercase letters, and digits. · Special characters are not supported.
Instance Name	Instance Name	The instance name. The name must be 2 to 128 characters in length. Spaces and special characters are not supported. Special characters include @/:="<>{[]}

4. Click Create.

   After creating the instance, wait until the instance status becomes Normal.

10.4.2 View instance details

After you create an instance, you can view the instance details in the KVStore for Memcache console.

Procedure

1. *Log on to the KVStore for Memcache console*.

2. In the instance list, click the instance ID or click the  icon in the Actions column and choose View Details from the shortcut menu. On the Instance Information page that appears, view the instance details.

   The Instance Information page contains the Basic Information, Configuration Information, and Connection Information sections. The *Instance information* table lists the configuration items in each section.

   Table 10-3: Instance Information page

Section	Item
Basic Information	· Instance ID · Name · Status · Region · Department · Project · Created At · Zone · Network Type · VPC (displayed only when the network type is VPC)
Configuration Information	· Instance Specification · Max Connections · Maximum Internal Network Bandwidth · Maintenance Time
Connection Information	· Connection Address · Port Number

10.4.3 Change the instance name

After you create an instance, you can change the name of the instance in the KVStore for Memcache console. This makes instances easy to locate by searching the specific instance name.

Procedure

1. *Log on to the KVStore for Memcache console*.

2. In the instance list, locate the instance, click  > Change in the Actions column.
3. In the Change Instance Information dialog box that appears, enter a new instance name in Name and click OK.

10.4.4 Change the instance specifications

KVStore for Memcache allows you to change the specifications of an instance.

Context

> **Note:**
> The instance will experience intermittent interruption for several seconds during specification change. We recommend that you change instance specifications during off-peak hours.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, click the instance ID or click the  icon in the Actions column and choose View Details from the shortcut menu. The Instance Information page appears.
3. Click Change Instance in the upper-right corner. In the Change Instance dialog box that appears, select the required instance specifications and click OK.
   The message `Instance changed.` is displayed. You need to wait until the instance status becomes Normal before using the instance.

10.4.5 Configure a whitelist

Before using an instance, you must set an IP address whitelist. For more information, see *Configure a whitelist*.

10.4.6 Configure a maintenance time period

You can specify a maintenance time period for an instance in the console. The instance is maintainable in the specified time period.

Context

To ensure the stability of KVStore for Memcache instances, the back-end system irregularly maintains instances and machines on the Apsara Stack platform.

Before the maintenance is performed, KVStore for Memcache sends SMS messages and emails to contacts listed in your Apsara Stack tenant account.

To guarantee the stability of the maintenance process, the instance will enter the Instance Maintaining state before the maintenance time. When the instance is in this state, access to data in the database is not affected. However, modification-related functions such as configuration changes are temporarily unavailable for this instance in the console, whereas query functions such as performance monitoring are still available.

Notice:
During maintenance, instances may experience intermittent interruption. We recommend that you set the maintenance time period to off-peak hours.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, click the Instance ID, or click the  icon in the Actions column and choose View Details from the shortcut menu. The Instance Information page appears.
3. Click Change Maintenance Window in the upper-right corner.

   The default maintenance time period for KVStore for Memcache is from 02:00 to 06:00.
4. Select a maintenance time period and click OK.

   The time period is in China Standard Time (UTC+8).

10.4.7 Clear the instance data

You can clear all data of an instance with a single click in the console. After the data is cleared, it cannot be restored.

Context

Note:
This operation will delete all data of an instance, and the data cannot be restored. Proceed with caution.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, find the instance that you want to clear, click the  icon in the Actions column, and choose Clear from the shortcut menu.
3. In the Clear Instance message that appears, click OK.

10.4.8 Reset the password

If you forget, need to change, or have not set a password for your instance, you can reset the password to create a new password.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, click the instance ID or click the  icon in the Actions column and choose View Details from the shortcut menu. The Instance Information page appears.
3. Click Reset Password in the upper-right corner.
4. In the Reset Password dialog box that appears, enter a new password and click Submit.

10.4.9 Parameter configuration

KVStore for Memcache supports six data eviction policies. You can modify `EvictionPolicy` in the Apsara Stack console to set an eviction policy that meets your business needs.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, click the instance ID or click the  icon in the Actions column and choose View Details from the shortcut menu. The Instance Information page appears.
3. Click the Parameters tab.

4. Click the icon in the Actions column corresponding to `EvictionPolicy` and choose  > Change from the shortcut menu.
5. Select a data eviction policy and click OK.

10.5 Backup and restore

10.5.1 Automatic backup

You can configure an automatic backup policy in the KVStore for Memcache console.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, click the Instance ID, or click  > View Details from the shortcut menu. The Instance Information page appears.
3. Click the Backup and Restore tab.
4. Click the Backup Settings tab.
5. Click Change Settings. In the Backup Settings dialog box that appears, configure Recurrence and Backup Time.

   Backup data is retained for seven days. You cannot modify this configuration.
6. Click OK to complete automatic backup configuration.

10.5.2 Manual backup

In addition to automatic backups, you can also manually back up data in the console.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, click the Instance ID, or click the  icon in the Actions column and choose View Details from the shortcut menu. The Instance Information page appears.
3. Click the Backup and Restore tab.
4. On the Backups tab, click Create Backup in the upper-right corner.

5. In the message that appears, click OK to back up the instance immediately.

   On the Backups tab, you can select the time range to query historical backup data. By default, backup data is retained for seven days, so you can query historical backup data from the past seven days.

10.5.3 Data restoration

The data restoration function minimizes losses caused by incorrect operations on database. KVStore for Memcache allows you to restore data from backup sets.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, click the instance ID or click the  icon in the Actions column and choose View Details from the shortcut menu. The Instance Information page appears.
3. Click the Backup and Restore tab.
4. On the Backup and Restore tab, click the Backups tab.
5. Find the backup file to be restored, click the  icon in the Actions column and choose Restore from the shortcut menu.
6. In the Restore message that appears, click OK to restore data from the backup file.

   Data restoration is considered a high-risk operation. Verify the data to be restored before performing this operation. Proceed with caution.

10.6 Supported protocols and commands

Any client compatible with the memcached protocol can access the KVStore for Memcache service. You can select any memcached client that is compatible with Simple Authentication and Security Layer (SASL) and the memcached binary protocol as you need.

Protocols

- The memcached binary protocol
- The SASL protocol

Commands

KVStore for Memcache supports the following commands.

Operation code	Operation command	Remarks
0x00	Get	-
0x01	Set	-
0x02	Add	-
0x03	Replace	-
0x04	Delete	-
0x05	Increment	-
0x06	Decrement	-
0x07	Quit	-
0x08	Flush	Memcache is accurate to the second.
0x09	GetQ	-
0x0a	No-op	-
0x0b	Version	-
0x0c	GetK	-
0x0d	GetKQ	-
0x0e	Append	-
0x0f	Prepend	-
0x10	Stat	Not supported
0x11	SetQ	-
0x12	AddQ	-
0x13	ReplaceQ	-
0x14	DeleteQ	-
0x15	IncrementQ	-
0x16	DecrementQ	-
0x17	QuitQ	-
0x18	FlushQ	-
0x19	AppendQ	-
0x1a	PrependQ	-

Operation code	Operation command	Remarks
0x1b	Verbosity	Not supported
0x1c	Touch	-
0x1d	GAT	-
0x1e	GATQ	-
0x20	SASL list mechs	-
0x21	SASL Auth	-
0x22	SASL Auth	-

11 ApsaraDB for MongoDB

11.1 What is ApsaraDB for MongoDB?

ApsaraDB for MongoDB is a stable, reliable, and scalable database service that fully complies with the MongoDB protocols. This service provides a full range of database solutions, including disaster recovery, backup, restore, monitoring, and alerts.

ApsaraDB for MongoDB provides the following features:

- **ApsaraDB for MongoDB automatically creates a three-node replica set and provides ready-to-use advanced features. These features include swithover and failover for disaster recovery. All the features are available to users.**
- **ApsaraDB for MongoDB allows you to back up and restore databases with ease . You can perform conventional database backup and database rollback in the ApsaraDB for MongoDB console.**
- **ApsaraDB for MongoDB supports alerts features and up to 20 performance metrics for you to view and monitor the database performance.**
- **ApsaraDB for MongoDB provides visual data management tools to facilitate database operations and maintenance.**

11.2 Instructions

You need to understand the precautions and restrictions of ApsaraDB for MongoDB before you start.

To ensure the stability and security of ApsaraDB for MongoDB instances, pay attention to the restrictions described in *Table 11-1: ApsaraDB for MongoDB restrictions*.

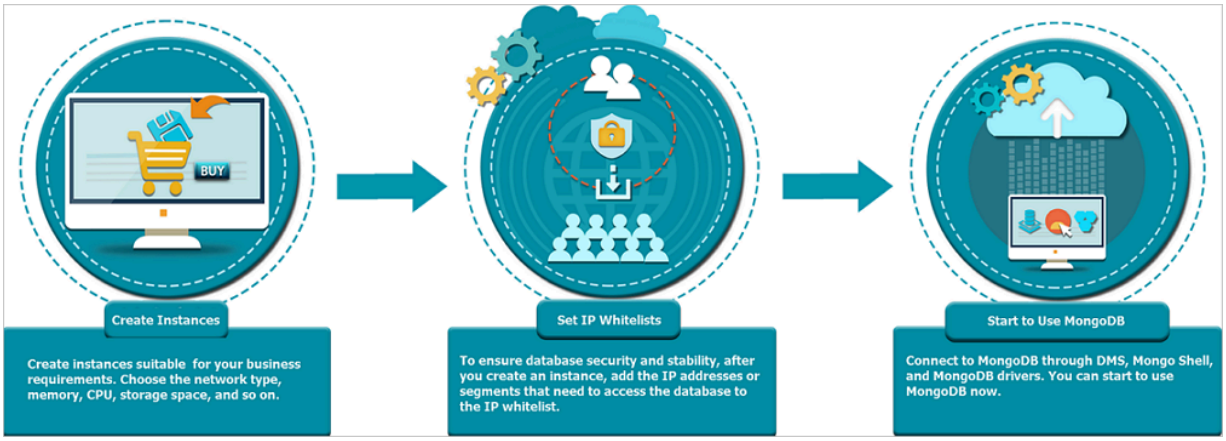
Table 11-1: ApsaraDB for MongoDB restrictions

Operation	Restrictions
Create a database copy	· The system automatically creates a three-node replica set. · The system provides the primary and secondary nodes. The standby node is hidden and invisible to the user. · The user cannot manually create the secondary nodes.
Restart a database	Instances must be restarted in the ApsaraDB for MongoDB console.

11.3 Quick start

11.3.1 Use ApsaraDB for MongoDB

This topic is a quick start guide to basic usage operations for ApsaraDB for MongoDB, such as creating an instance, configuring a whitelist, and connecting to an instance. Flowcharts are used to describe the basic procedures in ApsaraDB for MongoDB, and guide you to create an ApsaraDB for MongoDB instance.



- *Create an instance*

  An instance is a virtual database server on which you can create and manage multiple databases.

- *Set whitelist*

  After you create an ApsaraDB for MongoDB instance, you need to configure a whitelist for the instance to allow external devices access to the instance.

  A whitelist can enhance access security for ApsaraDB for MongoDB instances. We recommend that you update the whitelist regularly. Configuring the whitelist does not affect the normal services of the instance.

- *Use Mongo shell to connect to an instance*

  After you create an instance and configure a whitelist, you can use the mongo shell to connect to the instance.

11.3.2 Log on to the ApsaraDB for MongoDB console

This topic describes how to log on to the ApsaraDB for MongoDB console.

Prerequisites

- **Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://`IP address or domain name of the Apsara Stack console`/manage.**
- **We recommend that you use the Chrome browser.**

Procedure

1. **Open your browser.**
2. **In the address bar, enter the access address of the Apsara Stack console in the format of http://`IP address or domain name of the Apsara Stack console`/manage, and then press Enter.**
3. **Enter the correct username and password.**
   - **The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.**
   - **You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers**

(0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).

4. Click LOGIN to go to the Dashboard page.
5. In the top navigation bar, choose  > Database > ApsaraDB for MongoDB.

11.3.3 Create an instance

This topic describes how to create an instance in the ApsaraDB for MongoDB console.

Prerequisites

Make sure that you have created an account to log on to the ApsaraDB for MongoDB console.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click Create Instance in the upper-right corner of the page to go to the Create MongoDB Instance page. Follow the tips to configure parameters.

   The parameter configurations are described in *Table 11-2: Instance parameters*.

Table 11-2: Instance parameters

Category	Parameter	Description
Basic Settings	Department	The department to which an instance belongs.
	Project	The project to which an instance belongs.
	Region	Select a region for the instance.
	Zone	The available zone of the instance.

Category	Parameter	Description
Network Type	Network Type	Network types supported by ApsaraDB for MongoDB are as follows: · `Classic Network`: Cloud services on a classic network are not isolated. Unauthorized access to a cloud service is blocked only by the security group or whitelist policy of the service. · `VPC`: A virtual private cloud (VPC) helps you build an isolated network environment on Alibaba Cloud. You can customize the route table, IP address range, and gateway on a VPC. We recommend that you use VPC for improved security. You must create a VPC in advance. Or, you can change the network type after you create an instance.
Specifications	Node Specifications	ApsaraDB for MongoDB instance specifications are as follows: · Three-Member Replica Set: Uses dedicated memory and I/O resources while sharing CPU and storage resources with other general instances on the same server. · Exclusive Specifications: Uses dedicated CPU, memory, storage, and I/O resources to ensure long-term performance. This type of instances are not affected by other instances on the physical server. The maximum configuration for the exclusive specifications is Dedicated Hosts. All resources on the physical server are dedicated to this instance. · Dedicated Hosts
	Storage Space	The storage space of the instance, including the space for data, system files, binlog files, and transaction files.

Category	Parameter	Description
Password Settings	Set Password	The password used to log on to an ApsaraDB for MongoDB database. You can select Now to set the logon password immediately. Or, you can select Later and set the logon password later. For more information, see *Reset a password*. The password must meet the following requirements: · The password can contain lowercase letters, numbers, and underscores (_). · The password must be from 6 to 32 characters in length.
Instance Name	Instance name	It must start with an English letter or a Chinese character. It must be from 2 to 256 characters in length and can contain Chinese characters, English letters, numbers, underscores (_), and hyphens (-).

3. Click Create to create the instance.

11.3.4 Set whitelist

Before you use an ApsaraDB for MongoDB instance, you need to add IP addresses or IP segments used for database access to the whitelist of the instance to improve database security and stability. Proper use of the whitelist can enhance access security for ApsaraDB for MongoDB. We recommend that you maintain the whitelist regularly.

Context

The system creates a default whitelist group for each instance. This whitelist group can be modified or cleared, but cannot be deleted. After a new instance is created, the system automatically adds the IP address 0.0.0.0/0 to the default whitelist group. When the IP address 0.0.0.0/0 is on the whitelist, the instance is accessible from any IP address. To secure your database, delete the IP address 0.0.0.0/0 from the whitelist.

When the IP address 127.0.0.1 is on the whitelist, no IP addresses or IP address segments are allowed to access the instance. Make sure that the IP address 127.0.0.1 is not on the whitelist before you add any IP address or IP address segment.

Procedure

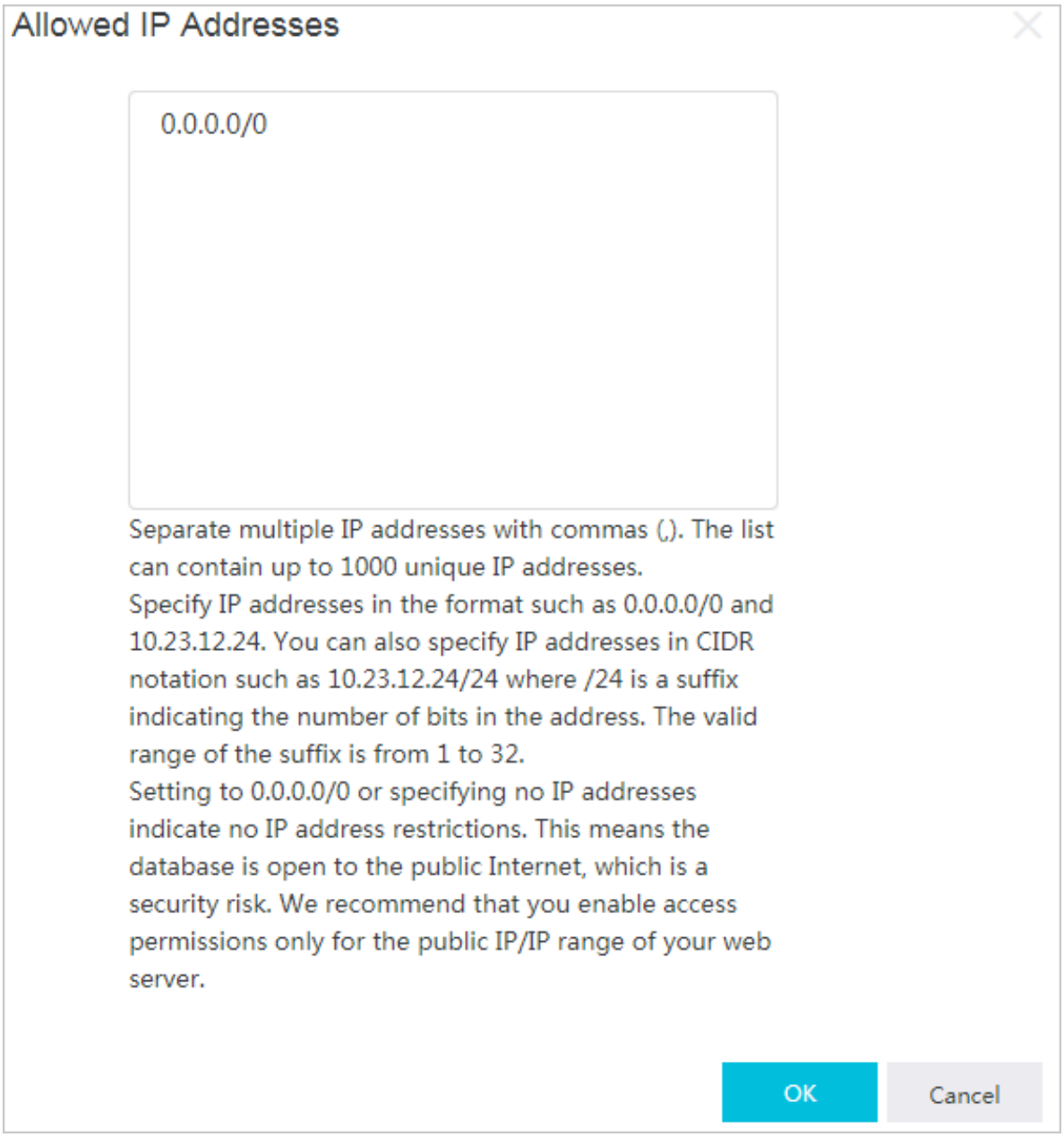
1. *Log on to the ApsaraDB for MongoDB console*.
2. **Click the ID of the instance to go to the Basic Information page.**
3. **In the left-side navigation pane, choose Security Control > Whitelist Settings to go to the Allowed IP Addresses page.**
4. **On the Allowed IP Addresses page, click Change Whitelist in the upper-right corner. In the Allowed IP Addresses dialog box that appears, follow the tips to configure parameters.**

   Enter IP addresses that are allowed to access the instance. Separate IP addresses with commas (,).

    **Note:**

Separate multiple IP addresses with commas (no space before or after each comma). For example, 192.168.0.1,172.16.213.9.

Figure 11-1: Allow access to IP addresses



5. **After you configure the parameters, click OK.**

11.3.5 Obtain the seven elements required to connect to an instance

ApsaraDB for MongoDB provides the connection addresses for two nodes in a three-node replica set. You can use these addresses to access the ApsaraDB for MongoDB

instance. This topic describes how to obtain the elements required to connect to an ApsaraDB for MongoDB instance.

Context

To access an ApsaraDB for MongoDB instance, obtain the following seven elements:

- Instance username
- Password
- Replica set name
- Domain names and port numbers of the two nodes

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the instance to go to the Basic Information page.
3. In the left-side navigation pane, click Database Connection. On the Network Information page that appears, six elements are displayed, as shown in *Figure 11-2: Network information*.

   Table 11-3: Element description describes the elements that are used to connect to an instance.

Figure 11-2: Network information

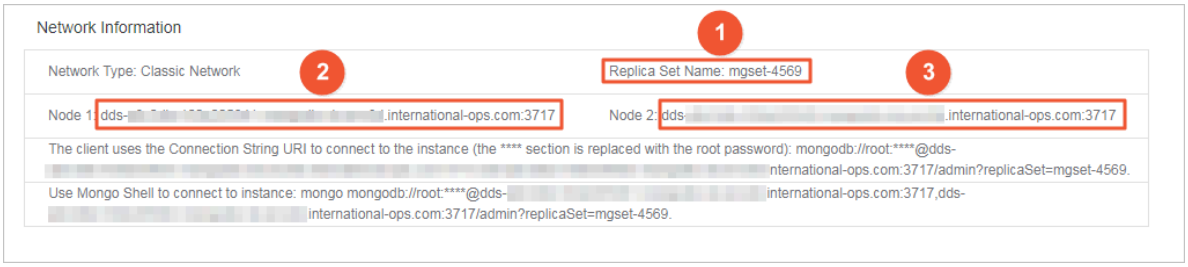


Table 11-3: Element description

Element	Description
Replica set name	Marked as 1 in the preceding figure.
Domain name of Node 1	Marked as 2 in the preceding figure.
Domain name of Node 2	Marked as 3 in the preceding figure.
Default account used for the initial logon to the database	The account is root.
Name of the default database	The database name is admin.

Element	Description
Database connection port	The database connection port is 3717.

Note:

The password for database connection is set when you create the instance. For more information about how to change this password, see *Reset a password*.

11.3.6 Use Mongo shell to connect to an instance

You can create an instance, configure the whitelist, and obtain the seven elements required for instance connection. This topic describes how to use Mongo shell to connect to an ApsaraDB for MongoDB instance.

Prerequisites

- **Before you use Mongo shell to connect to an ApsaraDB for MongoDB instance , you need to check that Mongo shell and the instance you want to connect to are deployed on the ECS instances in the same region and use the same type of networks.**
- **To ensure a successful authentication, use Mongo shell 3.0 or later versions.**

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. **Click the ID of the target instance to go to the Basic Information page.**
3. **In the left-side navigation pane, click Database Connection. The Network Information page displays six elements, including username, replica set name, and the domain name addresses and port numbers of the two nodes.**

   For more information about how to obtain these elements, see *Obtain the seven elements required to connect to an instance*.
4. **On the ECS, run the mongo command to connect to an ApsaraDB for MongoDB instance. Exampe:**

```
mongo --host dds-xxxx.mongodb.rds.aliyuncs.com:3717 -u root -p
123456 --authenticationDatabase admin
```

11.4 Instances

11.4.1 Create an instance

This topic describes how to create an instance in the ApsaraDB for MongoDB console.

Prerequisites

Make sure that you have created an account to log on to the ApsaraDB for MongoDB console.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. **Click Create Instance in the upper-right corner of the page to go to the Create MongoDB Instance page. Follow the tips to configure parameters.**

   The parameter configurations are described in *Table 11-4: Instance parameters*.

   Table 11-4: Instance parameters

Category	Parameter	Description
Basic Settings	Department	The department to which an instance belongs.
	Project	The project to which an instance belongs.
	Region	Select a region for the instance.
	Zone	The available zone of the instance.

Category	Parameter	Description
Network Type	Network Type	Network types supported by ApsaraDB for MongoDB are as follows: · `Classic Network`: Cloud services on a classic network are not isolated. Unauthorized access to a cloud service is blocked only by the security group or whitelist policy of the service. · `VPC`: A virtual private cloud (VPC) helps you build an isolated network environment on Alibaba Cloud. You can customize the route table, IP address range, and gateway on a VPC. We recommend that you use VPC for improved security. You must create a VPC in advance. Or, you can change the network type after you create an instance.
Specifications	Node Specifications	ApsaraDB for MongoDB instance specifications are as follows: · Three-Member Replica Set: Uses dedicated memory and I/O resources while sharing CPU and storage resources with other general instances on the same server. · Exclusive Specifications: Uses dedicated CPU, memory, storage, and I/O resources to ensure long-term performance. This type of instances are not affected by other instances on the physical server. The maximum configuration for the exclusive specifications is Dedicated Hosts. All resources on the physical server are dedicated to this instance. · Dedicated Hosts
	Storage Space	The storage space of the instance, including the space for data, system files, binlog files, and transaction files.

Category	Parameter	Description
Password Settings	Set Password	The password used to log on to an ApsaraDB for MongoDB database. You can select Now to set the logon password immediately. Or, you can select Later and set the logon password later. For more information, see *Reset a password*. The password must meet the following requirements: · The password can contain lowercase letters, numbers, and underscores (_). · The password must be from 6 to 32 characters in length.
Instance Name	Instance name	It must start with an English letter or a Chinese character. It must be from 2 to 256 characters in length and can contain Chinese characters, English letters, numbers, underscores (_), and hyphens (-).

3. Click Create to create the instance.

11.4.2 View instance details

You can view the details of an instance, such as the basic information, internal network connection information, running status, and configurations. This topic describes how to view instance details.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. You can go to the Instance Details page by using either of the following methods:
   - Click an instance ID to go the Basic Information page.
   - In the Actions column of the target instance, click > Query Details. On the Basic Information page, view basic information about the instance.

11.4.3 Restart an instance

You can manually restart an instance when the number of connections exceeds the threshold or any performance issue occurs on the instance. This topic describes how to restart an instance.

Context

> **Note:**
> **A restart will disconnect the instance. Make appropriate service arrangements before you restart an instance and perform this operation with caution.**

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. **In the Actions column of the target instance, click > Restart Instance.**
3. **In the Restart MongoDB Instance dialog box that appears, click OK.**

11.4.4 Change specifications

You can change the specifications of an instance, such as the memory and storage space, if the specifications are too high or too low to meet the performance requirements of an application.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. **In the Actions column of the target instance, click > Change Specifications.**
   In the Change MongoDB Specifications dialog box that appears, configure the parameters.
   When you change specifications, you can set Node Specifications and Storage Space of an instance.
3. **Specify the specifications and click OK.**

11.4.5 Switch to VPC

ApsaraDB for MongoDB supports classic networks and virtual private clouds (VPCs). You can switch between these two types of networks as required.

Context

The differences between a classic network and a VPC are as follows:

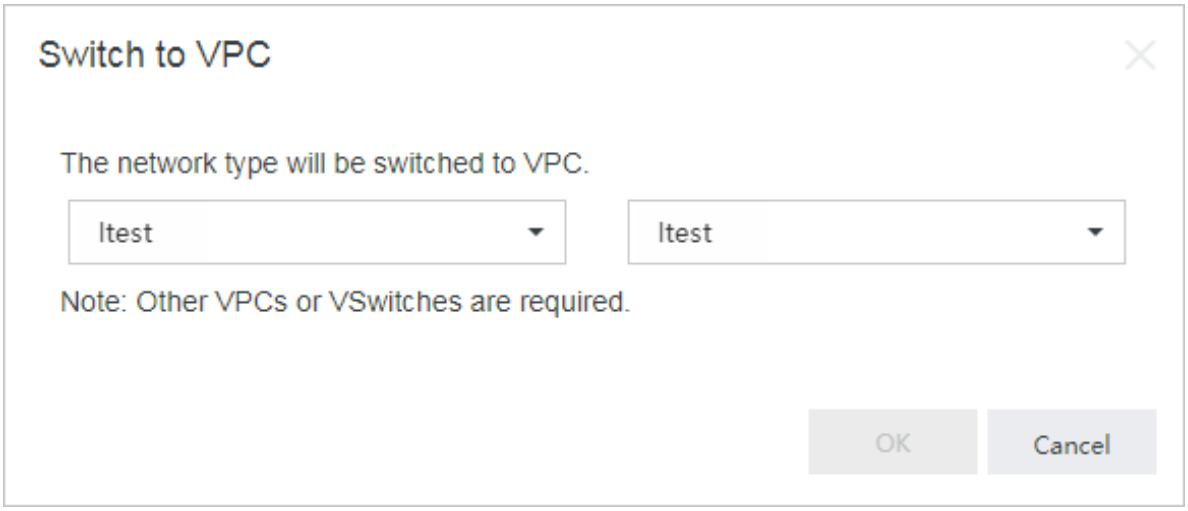
- **Classic network: The cloud service in a classic network is not isolated at the network layer. Unauthorized access is blocked only by the security group or whitelist policy of the cloud service.**
- **VPC: A VPC helps you build an isolated network environment on Alibaba Cloud . You can customize the route table, IP address range, and gateway for a VPC. In addition, you can use a leased line or VPN to combine your data center and the cloud resources in Apsara Stack to build a virtual data center and migrate your applications to the cloud.**

Note:
To use a VPC to create an ApsaraDB for MongoDB instance, make sure that the instance and VPC are in the same region.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. **Click the ID of the target instance to go to the Basic Information page.**
3. **In the left-side navigation pane, click Database Connection to go to the Network Information page.**
4. **On the Network Information page, click Switch to VPC. The Switch to VPC page appears.**

   Figure 11-3: Switch to VPC



5. **Specify a VPC and the associated VSwitch based on the description on the Switch to VPC page. Click OK.**

11.4.6 Modify an instance name

You can modify instance names to facilitate management. This topic describes how to modify an instance name.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. In the Actions column corresponding to the target instance, click . Select Change Instance Name. In the Change Instance Name page that appears, enter Instance Name.
3. In the `Instance Name` field, enter a new name for the instance.
4. Click OK.

   Note:

   - The instance name must start with an English letter or a Chinese character.
   - The instance name can contain Chinese characters, English letters, underscores (_), hyphens (-), and numbers.
   - It must be from 2 to 256 characters in length.
5. Click OK after you modify the instance name.

11.4.7 Reset a password

This topic describes how to reset your password in the ApsaraDB for MongoDB console.

Context

Notice:
For data security, we recommend that you change your password periodically.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the instance to go to the Basic Information page.

3. **On the Basic Information page, click Reset Password in the upper-right corner and configure parameters in the Reset Password dialog box that appears.**

   The parameter configurations are described in *Table 11-5: Password resetting parameters*.

   Table 11-5: Password resetting parameters

Parameter	Description
Logon Password	The password must be 6 to 32 characters in length and can contain letters, numbers, and underscores (_).
Confirm Password	The password must be 6 to 32 characters in length and can contain letters, numbers, and underscores (_).

4. **Click Submit.**

11.4.8 Release an instance

You can manually release an instance as needed. This topic describes how to manually release an instance.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. **Click the icon in the Actions column corresponding to the instance that you want to release, and click Delete Instance.**
3. **In the Delete MongoDB Instance dialog box that appears, click OK.**

11.5 Security

11.5.1 Set whitelist

Before you use an ApsaraDB for MongoDB instance, you need to add IP addresses or IP segments used for database access to the whitelist of the instance to improve database security and stability. Proper use of the whitelist can enhance access security for ApsaraDB for MongoDB. We recommend that you maintain the whitelist regularly.

Context

The system creates a default whitelist group for each instance. This whitelist group can be modified or cleared, but cannot be deleted. After a new instance is created,

the system automatically adds the IP address 0.0.0.0/0 to the default whitelist group. When the IP address 0.0.0.0/0 is on the whitelist, the instance is accessible from any IP address. To secure your database, delete the IP address 0.0.0.0/0 from the whitelist.

When the IP address 127.0.0.1 is on the whitelist, no IP addresses or IP address segments are allowed to access the instance. Make sure that the IP address 127.0.0.1 is not on the whitelist before you add any IP address or IP address segment.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. **Click the ID of the instance to go to the Basic Information page.**
3. **In the left-side navigation pane, choose Security Control > Whitelist Settings to go to the Allowed IP Addresses page.**
4. **On the Allowed IP Addresses page, click Change Whitelist in the upper-right corner. In the Allowed IP Addresses dialog box that appears, follow the tips to configure parameters.**

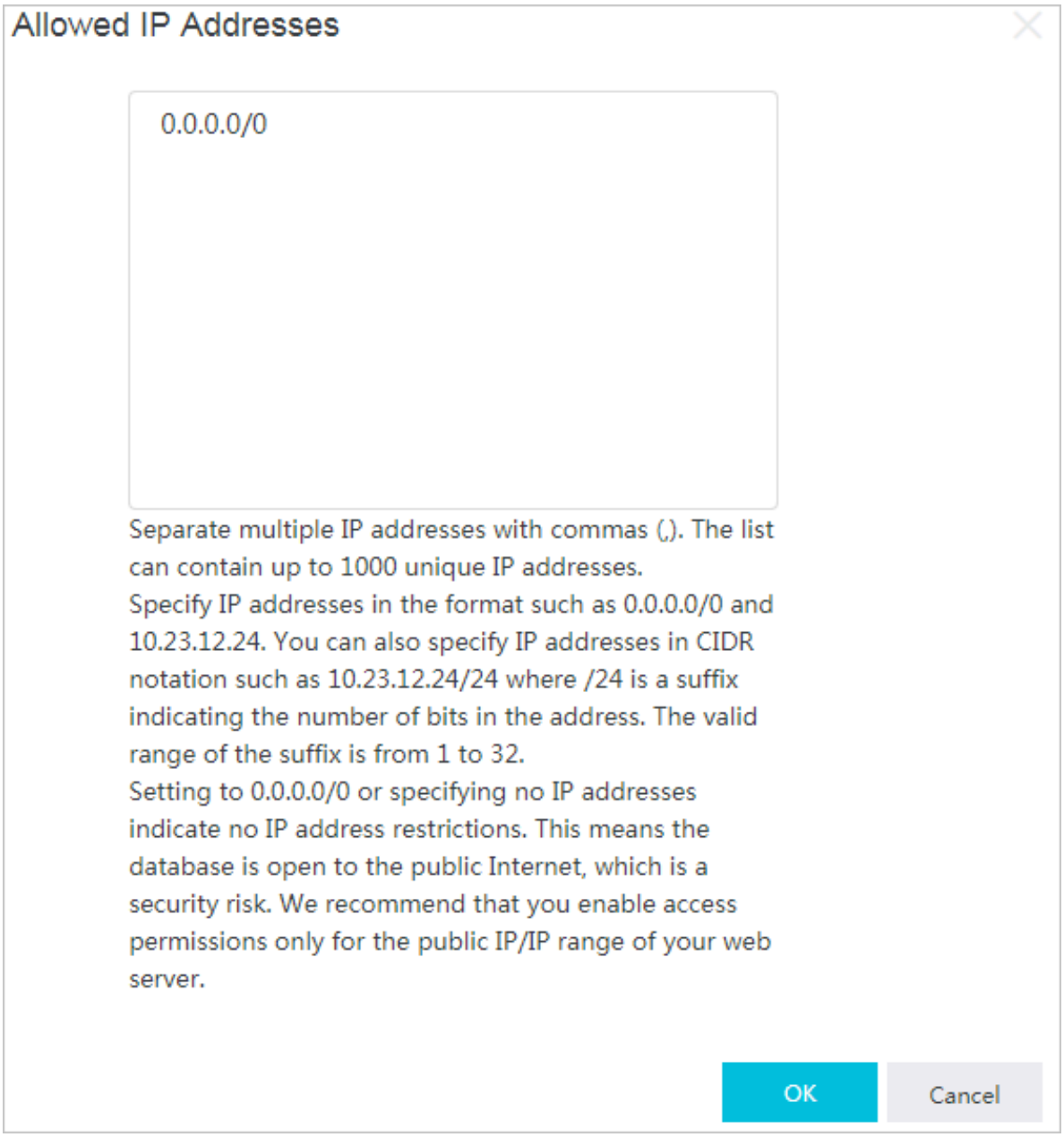
   Enter IP addresses that are allowed to access the instance. Separate IP addresses with commas (,).

   Note:

Separate multiple IP addresses with commas (no space before or after each comma). For example, 192.168.0.1,172.16.213.9.

Figure 11-4: Allow access to IP addresses



5. **After you configure the parameters, click OK.**

11.5.2 Audit logs

You can query the SQL logs, operation logs, and error logs of an instance in the ApsaraDB for MongoDB console to locate and analyze faults.

Context

The audit log service records all operations that a client performs on a connected database. This service provides references for fault analysis, behavior analysis, and security audit. The audit log service helps you obtain the data execution information for analysis. Audit logs are essential in the regulatory requirements of AntCloud and other core business scenarios.

> **Note:**
> **The storage cycle of audit logs is seven days. An audit log is automatically deleted seven days after it is generated.**

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. **Click the ID of the instance to go to the Basic Information page.**
3. **In the left-side navigation pane, choose Security Control > Audit Logs. The Audit Logs page appears.**
4. **On the Audit Logs page, you can search for and export logs.**
   - **Search: You can search an audit log based on the time range, database name, database account, or an execution statement.**
   - **Files: The list of audit log files.**
   - **Export File: Export audit log files.**

11.6 Monitoring information

The ApsaraDB for MongoDB console provides abundant performance metrics for you to check the running status of instances. You can check instance monitoring data in the ApsaraDB for MongoDB console.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. **Click the ID of the instance to go to the Basic Information page.**

3. **In the left-side navigation pane, click Monitoring Information.**

   You can select a time range to query historical metrics. Metric details are described in *Table 11-6: Metrics*.

   Table 11-6: Metrics

Metric	Description	Monitoring frequency	Monitoring period
CPU Utilization	The instance CPU utilization.	300 seconds/time	30 days
Memory Usage	The instance memory usage.	300 seconds/time	30 days
IOPS Usage	The IOPS used by the instance, including: · Data disk IOPS · Log disk IOPS	300 seconds/time	30 days
IOPS Usage	The percentage of the IOPS volume used by the instance to the maximum available IOPS volume.	300 seconds/time	30 days
Used Disk Space	The total disk space used by the instance, including: · Total usage · Data file usage · Log file usage	300 seconds/time	30 days
Disk Usage	The percentage of the total space used by the instance to the maximum available space.	300 seconds/time	30 days

Metric	Description	Monitoring frequency	Monitoring period
Opcounters	Operation QPS metrics on the instance, including: · The number of insert operations. · The number of query operations. · The number of delete operations. · The number of update operations. · The number of getmore operations. · The number of command operations.	300 seconds/time	30 days
Connections	The current number of connections to the instance.	300 seconds/time	30 days
Cursors	The number of cursors currently used by the instance, including: · The number of currently opened cursors. · The number of expired cursors.	300 seconds/time	30 days

Metric	Description	Monitoring frequency	Monitoring period
Network	The network traffic of the instance, including: · The inbound traffic. · The outbound traffic. · The number of processed requests.	300 seconds/time	30 days
GlobalLock	The length of the instance queue waiting for global lock, including: · The length of the instance queue waiting for global read lock. · The length of the instance queue waiting for global write lock. · The length of the instance queue waiting to perform operations on the global lock.	300 seconds/time	30 days

Metric	Description	Monitoring frequency	Monitoring period
WiredTiger	The cache indicators of the wiredTiger engine of an instance, including: · The volume of data read to the cache. · The capacity of the disk with data written from the cache. · The configured maximum available disk capacity.	300 seconds/time	30 days

11.7 Backup and restore data

11.7.1 Automatic backup

ApsaraDB for MongoDB allows you to specify backup settings and automatically backs up data based on the settings.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the instance to go to the Basic Information page.
3. In the left-side navigation pane, choose Backup and Restore > Backup Settings.

4. **Click the Backup Settings tab and click Change Settings. On the Backup Settings page that appears, specify the parameters.**

   The parameter configurations are described in *Table 11-7: Backup policy parameters*.

   Table 11-7: Backup policy parameters

Parameter	Description
Retention Period (Days)	Number of days for retaining data backups. The value range is from 1 to 30 days. Default value: 7 days.
Recurrence	One or multiple days in a week.
Backup Time	Any period of time in a day, in hours.

5. **After you complete the configuration, click Confirm.**

11.7.2 Back up instances manually

This topic describes how to manually back up an ApsaraDB for MongodDB instance.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. **Click the ID of the instance to go to the Basic Information page.**
3. **In the left-side navigation pane, choose Backup and Restore > Backup Management. The Backup and Restore page appears.**
4. **On the Backup and Restore page, click the Backups tab.**
5. **Click Back Up Instance. In the Back Up Instance dialog box that appears, click OK.**

11.7.3 Search for backups

This topic describes how to search for an instance backup.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. **Click the ID of the instance to go to the Basic Information page.**
3. **In the left-side navigation pane, choose Backup and Restore > Backup Management. The Backup and Restore page appears.**

4. On the Backup and Restore page, click the Backups tab, specify a time range, and click Search to search for the backups generated in the specified time range.

   Click in the leftmost column of a backup list to perform the following operations:

   - Download: Download the backup files that are generated in the specified time range. For more information, see *Download backups*.
   - Restore: Restore data from the backup files that are generated in the specified time range. For more information, see *Restore data*.
   - Create Instance from Backup: Create an instance from a specified backup. For information, see *Create an instance from a backup*. For information about how to create an instance, see *Create an instance*.

11.7.4 Restore data

The data restore function can minimize the damage caused by improper operations. ApsaraDB for MongoDB supports restoring data from a backup set.

Context

> **Note:**
> A rollback operation in ApsaraDB for MongoDB will overwrite data, and the overwritten data cannot be restored. Perform this operation with caution.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the instance to go to the Basic Information page.
3. In the left-side navigation pane, choose Backup and Restore > Backup Management. The Backup and Restore page appears.
4. Click the Backups tab.
5. In the Actions column corresponding to the target backup list, click > Restore. In the Restore dialog box that appears, click OK.

11.7.5 Download backups

ApsaraDB for MongoDB allows you to download backup files in the ApsaraDB for MongoDB console.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the instance to go to the Basic Information page.
3. In the left-side navigation pane, choose Backup and Restore > Backup Management. The Backup and Restore page appears.
4. Click the Backups tab.
5. In the Actions column of the target backup list, click > Download.
6. In the Download dialog box that appears, click Confirm to download the backup file to a local disk.

11.7.6 Create an instance from a backup

You can use a backup file to create a new instance as needed. The new instance contains all the data in the backup set.

Context

> **Note:**
> The storage space of the new instance must be equal to or greater than that of the source instance.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the instance to go to the Basic Information page.
3. In the left-side navigation pane, choose Backup and Restore > Backup Management. The Backup and Restore page appears.
4. Click the Backups tab.
5. In the Actions column corresponding to the target backup list, click > Create Instance from Backup. On the Create Instance from Backup page, click OK to go to the Create MongoDB Instance page.

   For more information about how to create an instance, see *Create an instance*.

12 AnalyticDB for PostgreSQL

12.1 What is AnalyticDB for PostgreSQL?

AnalyticDB for PostgreSQL (formerly known as HybridDB for PostgreSQL) is a distributed cloud database that uses multiple compute groups to provide Massively Parallel Processing (MPP) data warehousing service.

AnalyticDB for PostgreSQL is developed based on the Greenplum Open Source Database project and has been enhanced by Alibaba Cloud. This service has the following features:

- **Compatible with Greenplum, allowing you to use all tools that support Greenplum.**
- **Supports OSS, JSON, and HyperLogLog, a probability cardinality estimation algorithm.**
- **Supports flexible hybrid analysis by using the standard query syntax of SQL 2008 and OLAP aggregate functions.**
- **Provides a hybrid mode that supports both column store and row store, enhancing analytics performance.**
- **Supports data compression technology that can reduce storage costs.**
- **Provides online expansion and performance monitoring services to free you from managing and maintaining large numbers of MPP clusters. This enables DBAs, developers, and data analysts to focus on improving enterprise productivi ty and creating core business by using SQL.**

12.2 Quick start

12.2.1 Overview

This topic provides a quick start guide about how to manage AnalyticDB for PostgreSQL instances, such as creating an instance and logging on to a database.

- *Log on to the AnalyticDB for PostgreSQL console*

  This topic describes how to log on to the AnalyticDB for PostgreSQL console.

- *Create an instance*

  You can create an instance in the console and then manage the instance.

- *Configure a whitelist*

  To ensure a secure and stable database, before you use an AnalyticDB for PostgreSQL instance, you need to add IP addresses or CIDR blocks that are allowed to access the database to a whitelist of the instance.

- *Create an initial account*

  After you create an instance, you must create an initial account to log on to the database.

- *Connect to the database*

  You can use a client that supports PostgreSQL or Greenplum to connect to the database.

12.2.2 Log on to the AnalyticDB for PostgreSQL console

This topic describes how to log on to the AnalyticDB for PostgreSQL console.

Prerequisites

- **Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://`IP address or domain name of the Apsara Stack console`/manage.**
- **We recommend that you use the Chrome browser.**

Procedure

1. **Open your browser.**
2. **In the address bar, enter the access address of the Apsara Stack console in the format of http://`IP address or domain name of the Apsara Stack console`/manage, and then press Enter.**
3. **Enter the correct username and password.**
   - **The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.**
   - **You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the**

password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).

4. Click LOGIN to go to the Dashboard page.
5. In the top navigation bar, choose  > Database > AnalyticDB for PostgreSQL.

12.2.3 Create an instance

You can create an instance in the console and then manage the instance.

Prerequisites

You must create a VPC and a VSwitch.

Procedure

1. *Log on to the AnalyticDB for PostgreSQL console*.
2. On the AnalyticDB for PostgreSQL page, click Create Instance in the upper-right corner of the page.
3. On the Create Instance page, configure the following parameters. *Table 12-1: Instance creation parameters* describes the parameters.

Table 12-1: Instance creation parameters

Section	Parameter	Description
Region	Region	The region of the instance. If you need to access the AnalyticDB for PostgreSQL instance from an ECS instance over VPC, you must deploy the instance in the same region and zone as those of the ECS instance.  Note: You can choose from multiple regions.
Basic Settings	Department	The department to which the instance belongs.
	Project	The project to which the instance belongs.
	Zone	Select a zone where the instance is located.

Section	Parameter	Description
	Engine	Currently, only the integrated computing and storage version is supported.
	Computing Group Specifications	The unit of computing resources. Different group types have different storage capacities and computing capabilities.
	Computing Group Nodes	The number of `compute groups`. An instance must contain at least two compute groups. The performance of an instance scales linearly with the number of compute groups.
Network	Network Type	VPC (Virtual Private Cloud): You can use a VPC to build an isolated network environment in Alibaba Cloud. You can customize the routing table, IP address range, and gateway in a VPC. **Note:** Make sure you create a VPC and a VSwitch before you create an instance. For more information, see the "Create a default VPC and VSwitch" section in VPC User Guide.

4. Click Create.

Note:
An AnalyticDB for PostgreSQL database takes some time to initialize. You can perform operations when the instance is in the `running` state.

12.2.4 Configure a whitelist

To ensure a secure and stable database, you need to add IP addresses or CIDR blocks that are allowed to access the database to a whitelist.

Procedure

1. *Log on to the AnalyticDB for PostgreSQL console*.
2. Click the ID of the instance or click the icon in the Actions column and choose Details from the shortcut menu.

3. In the left-side navigation pane, click Data Security to go to the Data Security page.
4. You can use the following two methods to configure a whitelist.
   - Configure the default whitelist: Click the  icon corresponding to the default whitelist to go to the Change Whitelist Group page.
   - Create a custom whitelist: Click Create Whitelist Group.

Table 12-2: Whitelist configuration parameters

Parameter	Description
Group Name	Specify the name of the whitelist. The whitelist name must be 2 to 32 characters in length and can contain lowercase letters, digits, and underscores (_). It must start with a lowercase letter and end with a letter or digit. The default whitelist cannot be modified or deleted.
IP Addresses	Enter the CIDR blocks or IP addresses that are allowed to access the database. Use commas (,) to separate multiple CIDR blocks or IP addresses. • The whitelist contains the IP address 127.0.0.1 by default, indicating that no external IP addresses are allowed to access the instance. You must delete 127.0.0.1 before adding other IP addresses or CIDR blocks to the whitelist. • A whitelist can contain IP addresses such as 10.10.10.1 and CIDR blocks such as 10.10.10.0/24. This CIDR block indicates that any IP addresses in the 10.10.10.X format have access to the database. • The percent sign (%) or 0.0.0.0/0 indicates that any IP addresses are allowed to access the database.  **Warning:** This configuration is not recommended because it reduces the security of the database.

5. After you complete the configuration, click OK.

What's next

We recommend that you maintain the whitelist on a regular basis to ensure security of AnalyticDB for PostgreSQL.

You can click the icon corresponding to the whitelist to modify the whitelist.

12.2.5 Create an initial account

After you create an instance, you must create an initial account to log on to the database.

Procedure

1. *Log on to the AnalyticDB for PostgreSQL console*.
2. Click the ID of the instance, or click the icon in the Actions column and choose Details from the shortcut menu.
3. In the left-side navigation pane, click Accounts to go to the Accounts page.
4. Click Create Initial Account to go to the create account page.
5. On the Create Account page, enter the account name and password.

   Table 12-3: Parameter description

Parameter	Description
Database Account	The name of the account must be 2 to 16 characters in length and can contain lowercase letters, digits, and underscores (_). It must start with a letter and end with a letter or digit, for example, *user4example*.
Password	The password must be 8 to 32 characters in length. It must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters.

6. Click OK.

> **Note:**
> After the initial account is created, you cannot use the console to delete the account or add other accounts. However, you can execute SQL statements in the database to create other accounts.

12.2.6 Obtain the client tool

The interface protocol of AnalyticDB for PostgreSQL is compatible with Greenplum Community Edition and PostgreSQL 8.2. Therefore, you can use the Greenplum or PostgreSQL client to connect to AnalyticDB for PostgreSQL.

> **Note:**
>
> **Apsara Stack is an isolated environment. You must deploy the software installation packages to the internal environment.**

Graphical client tools

AnalyticDB for PostgreSQL users can directly use client tools that support Greenplum, such as *SQL Workbench*, *Navicat Premium*, *Navicat for PostgreSQL*, and *pgAdmin III (1.6.3)*.

Command-line client psql (for RHEL 6, RHEL 7, CentOS 6, and CentOS 7)

For Red Hat Enterprise Linux (RHEL) and CentOS 6 or 7, you can download the tool from the following addresses and decompress the package to use it:

- **For RHEL version 6 or CentOS version 6, click *hybriddb_client_package_el6*.**
- **For RHEL version 7 or CentOS version 7, click *hybriddb_client_package_el7*.**

Command-line client psql (for other Linux systems)

The compilation methods for client tools applicable to other Linux systems are as follows:

1. **Obtain the source code by using one of the following methods:**
   - **Obtain the git directory (you need to install the git tool first).**

     ```
     git clone https://github.com/greenplum-db/gpdb.git
     cd gpdb
     git checkout 5d870156
     ```

   - **Download the code.**

     ```
     wget https://github.com/greenplum-db/gpdb/archive/5d87015609
     abd330c68a5402c1267fc86cbc9e1f.zip
     unzip 5d87015609abd330c68a5402c1267fc86cbc9e1f.zip
     cd gpdb-5d87015609abd330c68a5402c1267fc86cbc9e1f
     ```

2. **Use gcc and other compilation tools.**

   ```
   ./configure
   make -j32
   ```

```
make install
```

3. **Use psql and pg_dump. The paths of the two tools are as follows:**

   psql: `/usr/local/pgsql/bin/psql`

   pg_dump: `/usr/local/pgsql/bin/pg_dump`

Command-line client psql (for Windows and other systems)

For the client tools for Windows and other systems, go to the Pivotal website to download *HybridDB Client*.

12.2.7 Connect to the database

The Greenplum Database is developed based on the PostgreSQL 8.2 branch and fully compatible with its message protocol. AnalyticDB for PostgreSQL is also based on the same PostgreSQL version. Therefore, AnalyticDB for PostgreSQL users can use tools that support the PostgreSQL 8.2 message protocol, such as libpq, JDBC, ODBC, psycopg2, and pgAdmin III.

Context

AnalyticDB for PostgreSQL provides a binary psql client for Red Hat Enterprise Linux. To download the program, see *Obtain the client tool*. The Greenplum official website provides an installation package, which includes JDBC, ODBC, and libpq. The package is easy to install and use. For more information, see *Greenplum documentation*.

Note:

- **Apsara Stack is an isolated environment. To access Apsara Stack, you need to prepare the necessary software installation packages in advance.**
- **By default, you can only use clients deployed in the ECS instance that is in the same region and zone as the PostgreSQL database to access AnalyticDB for PostgreSQL.**

psql

psql is a common tool used together with Greenplum, and provides a variety of command functions. Its binary files are located in the `bin` directory of Greenplum. The procedure is as follows:

1. **Connect to AnalyticDB for PostgreSQL by using one of the following methods:**

   - **Connection string**

     ```
     psql "host=yourgpdbaddress.gpdb.rds.aliyuncs.com port=3432 dbname=
     postgres user=gpdbaccount password=gpdbpassword"
     ```

   - **Specified parameters**

     ```
     psql -h yourgpdbaddress.gp.aliyun-inc.com -p 3432 -d postgres -U
     gpdbaccount
     ```

     Parameters:

     - **-h: specifies the host address.**
     - **-p: specifies the port number.**
     - **-d: specifies the database. The default database is postgres.**
     - **-U: specifies the user to connect to the database.**

     In psql, you can run the `psql --help` command to view more options. You can run the `\?` command to view the commands supported in psql.

2. **Enter the password to go to the psql shell interface.**

   ```
   postgres=>
   ```

Reference

- **AnalyticDB for PostgreSQL also supports the psql commands of PostgreSQL. Pay attention to the details of the differences. For more information, see *PostgreSQL 8.3 .23 Documentation - psql*.**

pgAdmin III

PgAdmin III is a graphical client for PostgreSQL and can be used to connect directly to AnalyticDB for PostgreSQL. For more information, click *here*. For more information about other graphics clients, see *Obtain the client tool*.

1. **Download pgAdmin III 1.6.3 or earlier versions.**

   You can download pgAdmin III 1.6.3 from the *PostgreSQL website*. PgAdmin III 1.6.3 supports various operating systems, such as Windows, MacOS, and Linux.

    **Note:**

AnalyticDB for PostgreSQL is compatible with PostgreSQL 8.2. Therefore, you must use pgAdmin III 1.6.3 or an earlier version to connect to AnalyticDB for PostgreSQL.

2. **Choose File > Add Server.**
3. **In the New Server Registration dialog box that appears, enter the configuration information.**
4. **Click OK to connect to AnalyticDB for PostgreSQL.**

JDBC

JDBC uses the interface provided by PostgreSQL. The download methods are as follows:

- **Click *PostgreSQL JDBC Driver* to download the official JDBC of PostgreSQL, and then add it to the environment variables.**

Sample code

```
import java.sql.Connection; import java.sql.DriverManager; import java.
sql.ResultSet; import java.sql.SQLException; import java.sql.Statement
; public class gp_conn { public static void main(String[] args) { try
 { Class.forName("org.postgresql.Driver"); Connection db = DriverMana
ger.getConnection("jdbc:postgresql://mygpdbpub.gpdb.rds.aliyuncs.com:
3432/postgres","mygpdb","mygpdb"); Statement st = db.createStatement();
ResultSet rs = st.executeQuery("select * from gp_segment_configuration
;"); while (rs.next()) { System.out.print(rs.getString(1)); System.out
.print(" | "); System.out.print(rs.getString(2)); System.out.print(" |
 "); System.out.print(rs.getString(3)); System.out.print(" | "); System
.out.print(rs.getString(4)); System.out.print(" | "); System.out.print(
rs.getString(5)); System.out.print(" | "); System.out.print(rs.getString
(6)); System.out.print(" | "); System.out.print(rs.getString(7)); System
.out.print(" | "); System.out.print(rs.getString(8)); System.out.print
(" | "); System.out.print(rs.getString(9)); System.out.print(" | ");
System.out.print(rs.getString(10)); System.out.print(" | "); System.out.
println(rs.getString(11)); } rs.close(); st.close(); } catch (ClassNotFo
undException e) { e.printStackTrace(); } catch (SQLException e) { e.
printStackTrace(); } } }
```

Python

Python uses psycopg2 to connect to Greenplum and PostgreSQL. The procedure is as follows:

1. **Install psycopg2. There are three installation methods in CentOS:**
   - **Method 1: Run the `yum -y install python-psycopg2` command.**
   - **Method 2: Run the `pip install psycopg2` command.**
   - **Method 3: Run the source code:**

     ```
     yum install -y postgresql-devel*
     ```

```
wget http://initd.org/psycopg/tarballs/PSYCOPG-2-6/psycopg2-2.6.
tar.gz
tar xf psycopg2-2.6.tar.gz
cd psycopg2-2.6
python setup.py build
sudo python setup.py install
```

2. **Run the following commands to set PYTHONPATH and reference it:**

```
import psycopg2
 sql = 'select * from gp_segment_configuration;'
 conn = psycopg2.connect(database='gpdb', user='mygpdb', password='
mygpdb', host='mygpdbpub.gpdb.rds.aliyuncs.com', port=3432)
 conn.autocommit = True
 cursor = conn.cursor()
 cursor.execute(sql)
 rows = cursor.fetchall()
 for row in rows:
     print row
 conn.commit()
 conn.close()
```

A similar output is displayed:

```
(1, -1, 'p', 'p', 's', 'u', 3022, '192.168.2.158', '192.168.2.158',
None, None)(6, -1, 'm', 'm', 's', 'u', 3019, '192.168.2.47', '192.168
.2.47', None, None)(2, 0, 'p', 'p', 's', 'u', 3025, '192.168.2.148',
 '192.168.2.148', 3525, None)(4, 0, 'm', 'm', 's', 'u', 3024, '192.168
.2.158', '192.168.2.158', 3524, None)(3, 1, 'p', 'p', 's', 'u', 3023,
 '192.168.2.158', '192.168.2.158', 3523, None)(5, 1, 'm', 'm', 's', 'u
', 3026, '192.168.2.148', '192.168.2.148', 3526, None)
```

libpq

libpq is the C language interface to AnalyticDB for PostgreSQL. You can use the libpq library to access and manage PostgreSQL databases in a C program. You can locate its static and dynamic libraries under the lib directory.

For the example programs, click *Example Programs*.

For more information about libpq, see *PostgreSQL 9.4.17 Documentation - Chapter 31. libpq - C Library*.

ODBC

PostgreSQL ODBC is an open-source version based on the LGPL (GNU Lesser General Public License) protocol. You can download it from the *PostgreSQL website*.

1. **Install the driver.**

```
yum install -y unixODBC.x86_64
```

```
yum install -y postgresql-odbc.x86_64
```

2. **View the driver configuration.**

```
cat /etc/odbcinst.ini
# Example driver definitions
# Driver from the postgresql-odbc package
# Setup from the unixODBC package
[PostgreSQL]
Description = ODBC for PostgreSQL
Driver = /usr/lib/psqlodbcw.so
Setup = /usr/lib/libodbcpsqlS.so
Driver64 = /usr/lib64/psqlodbcw.so
Setup64 = /usr/lib64/libodbcpsqlS.so
FileUsage = 1
# Driver from the mysql-connector-odbc package
# Setup from the unixODBC package
[MySQL]
Description = ODBC for MySQL
Driver = /usr/lib/libmyodbc5.so
Setup = /usr/lib/libodbcmyS.so
Driver64 = /usr/lib64/libmyodbc5.so
Setup64 = /usr/lib64/libodbcmyS.so
FileUsage = 1
```

3. **Configure the DSN. Replace the `****` in the following code with the corresponding connection information.**

```
[mygpdb]
Description = Test to gp
Driver = PostgreSQL
Database = ****
Servername = ****.gpdb.rds.aliyuncs.com
UserName = ****
Password = ****
Port = ****
ReadOnly = 0
```

4. **Test connectivity.**

```
echo "select count(*) from pg_class" | isql mygpdb
+---------------------------------------+
| Connected!                            |
|                                       |
| sql-statement                         |
| help [tablename]                      |
| quit                                  |
|                                       |
+---------------------------------------+
SQL> select count(*) from pg_class
+----------------------+
| count                |
+----------------------+
| 388                  |
+----------------------+
SQLRowCount returns 1
1 rows fetched
```

5. **After ODBC is connected to the instance, connect the application to ODBC. For more information, see** *PostgreSQL ODBC Driver* **and** *psqlODBC HOWTO - C#*.

Reference

- *Pivotal Greenplum documentation*
- *PostgreSQL psqlODBC*
- *Compiling psqlODBC on Unix*
- *Download ODBC connectors*
- *Download JDBC connectors*
- *The PostgreSQL JDBC Interface*

12.3 Instances

12.3.1 Reset the password

If you forget the password of your database account, you can reset the password in the AnalyticDB for PostgreSQL console.

> **Notice:**
> **To ensure data security, we recommend that you change your password periodically.**

1. *Log on to the AnalyticDB for PostgreSQL console*.
2. **Click the ID of the instance, or click the icon in the Actions column and choose Details from the shortcut menu.**
3. **In the left-side navigation pane, click Accounts to go to the Accounts page.**
4. **Click the icon and choose Reset Password to go to the Reset password page.**
5. **After you enter and confirm the new password, click OK.**

12.3.2 View monitoring information

You can go to the monitoring information page in the console to view the operation status of an instance.

Procedure

1. *Log on to the AnalyticDB for PostgreSQL console*.
2. **Click the ID of the instance or click the icon in the Actions column and choose Details from the shortcut menu.**

3. In the left-side navigation pane, click Monitoring Information to go to the Monitoring information page.

   Specify a period of time, n, to view the metrics of the last n period of time. The maximum period of time is one day.

12.3.3 Switch the network type of an instance

The default network type of an instance is VPC. After you create an instance, you can switch between classic network and VPC as needed.

Context

AnalyticDB for PostgreSQL supports two network types: classic network and Virtual Private Cloud (VPC). Both network types use BGP connections, and are independent of the public network of your service provider. These network types only differ in function, which you can choose based on your requirements. The two network types are applicable to different scenarios:

- **Classic network: IP addresses are allocated by Alibaba Cloud. Classic networks are easy to configure and use. This network type is suitable for users who do not need to perform complex operations, or require short deployment cycles.**
- **VPC: VPCs are logically isolated private networks that support leased line connections. You can customize the network topology and IP addresses. This network type is suitable for advanced users.**

> **Warning:**
> **Switching the network type will cause the database service to stop. Proceed with caution.**

The following operations describe how to switch from a VPC to a classic network.

Procedure

1. *Log on to the AnalyticDB for PostgreSQL console*.
2. In the instance list, click the icon in the Actions column and choose Details from the shortcut menu to go to the Basic Information page.
3. Click the Database Connection tab to go to the database connection page.

4. Click Switch to Classic Network on the page. In the message that appears, click OK to switch the network type.

> **Note:**
>
> - After you switch the network type, it takes 3 to 30 minutes for the instance to enter the running state.
> - Before you switch from a classic network to a VPC, you must create a VPC and a VSwitch. During the operation, you must select a VPC and a VSwitch. You can also configure private IP addresses as needed.

12.3.4 Restart an instance

Each time a new version of AnalyticDB for PostgreSQL is released, you can restart an instance to update the database kernel version, allowing you to use the extended features in the new version.

Context

> **Warning:**
> Restarting an instance will cause the database service to stop. Proceed with caution.

Procedure

1. *Log on to the AnalyticDB for PostgreSQL console*.
2. In the instance list, click the icon in the Actions column and choose Restart from the shortcut menu. In the message that appears, click OK.

   After you initiate a restart, the instance enters the `Restarting` state. After the restart is complete, the instance enters the `Running` state.

   The restart process typically takes 3 to 30 minutes. During the restart period , the instance cannot provide external services. We recommend that you take precautionary measures before restarting instances. After the instance is restarted and is in the running state, you can access the database.

12.3.5 Import Data

12.3.5.1 High-speed parallel import of OSS

AnalyticDB for PostgreSQL can import or export data from or to OSS tables in parallel by using the OSS external table function, gpossext. AnalyticDB for PostgreSQL also supports GZIP compression for the OSS external tables to reduce file size and storage costs. gpossext can read from and write to TEXT and CSV files as well as GZIP compressed TEXT and CSV files.

- **Create an OSS external table extension (oss_ext)**

  To use an OSS external table, you must first create an OSS external table extension in AnalyticDB for PostgreSQL. You must create an extension for each database that you need to access.

  - **Creation statement:** `CREATE EXTENSION IF NOT EXISTS oss_ext;`
  - **Deletion statement:** `DROP EXTENSION IF EXISTS oss_ext;`

- **Import data in parallel**

  1. **Distribute data evenly among multiple OSS files for storage. We recommend that you set the number of OSS files to an integer that is the multiple of the number of segments in AnalyticDB for PostgreSQL.**
  2. **Create a READABLE external table in AnalyticDB for PostgreSQL.**
  3. **Execute the following statement to import data in parallel:**

     ```
     INSERT INTO <destination table> SELECT * FROM <external table>
     ```

  Note:

  - **The data import performance depends on the OSS performance as well as resources of the AnalyticDB for PostgreSQL cluster, such as CPU, I/O, memory, and network resources. We recommend that you use column store and compression when creating a table to achieve the best import performance. For example, you can specify the following clause:** `WITH (APPENDONLY=true, ORIENTATION=column, COMPRESSTYPE=zlib, COMPRESSLEVEL=5, BLOCKSIZE=1048576)`**. For more information, see** *Greenplum Database official documentation on database table creation syntax*.

- **Export data in parallel**

  1. **Create a WRITABLE external table in AnalyticDB for PostgreSQL.**
  2. **Execute the following statement to export data to OSS in parallel:**

```
INSERT INTO <external table> SELECT * FROM <source table>
```

- **Create OSS external tables**

Note:

The syntax for creating and using external tables is the same as that of Greenplum Database except for the syntax of location-related parameters.

```
CREATE [READABLE] EXTERNAL TABLE tablename
( columnname datatype [, ...] | LIKE othertable )
LOCATION ('ossprotocol')
FORMAT 'TEXT'
            [( [HEADER]
               [DELIMITER [AS] 'delimiter' | 'OFF']
               [NULL [AS] 'null string']
               [ESCAPE [AS] 'escape' | 'OFF']
               [NEWLINE [ AS ] 'LF' | 'CR' | 'CRLF']
               [FILL MISSING FIELDS] )]
          | 'CSV'
            [( [HEADER]
               [QUOTE [AS] 'quote']
               [DELIMITER [AS] 'delimiter']
               [NULL [AS] 'null string']
               [FORCE NOT NULL column [, ...]]
               [ESCAPE [AS] 'escape']
               [NEWLINE [ AS ] 'LF' | 'CR' | 'CRLF']
               [FILL MISSING FIELDS] )]
[ ENCODING 'encoding' ]
[ [LOG ERRORS [INTO error_table]] SEGMENT REJECT LIMIT count
       [ROWS | PERCENT] ]
CREATE WRITABLE EXTERNAL TABLE table_name
( column_name data_type [, ...] | LIKE other_table )
LOCATION ('ossprotocol')
FORMAT 'TEXT'
               [( [DELIMITER [AS] 'delimiter']
               [NULL [AS] 'null string']
               [ESCAPE [AS] 'escape' | 'OFF'] )]
          | 'CSV'
               [([QUOTE [AS] 'quote']
               [DELIMITER [AS] 'delimiter']
               [NULL [AS] 'null string']
               [FORCE QUOTE column [, ...]] ]
               [ESCAPE [AS] 'escape'] )]
[ ENCODING 'encoding' ]
[ DISTRIBUTED BY (column, [ ... ] ) | DISTRIBUTED RANDOMLY ]
ossprotocol:
   oss://oss_endpoint prefix=prefix_name
    id=userossid key=userosskey bucket=ossbucket compressiontype=[
none|gzip] async=[true|false]
ossprotocol:
   oss://oss_endpoint dir=[folder/[folder/]...]/file_name
    id=userossid key=userosskey bucket=ossbucket compressiontype=[
none|gzip] async=[true|false]
```

```
ossprotocol:
   oss://oss_endpoint filepath=[folder/[folder/]...]/file_name
    id=userossid key=userosskey bucket=ossbucket compressiontype=[
none|gzip] async=[true|false]
```

Parameters

Table 12-4: Common parameters

Parameter	Description
Protocol and endpoint	**It is in the protocol name://oss_endpoint format. The protocol name is oss. oss_endpoint is the domain name used by users to access OSS in a region.**  **Note:** **You can access the database from a VPC host by using an internal endpoint containing "internal" in the name in order not to generate public traffic.**
id	**The ID of the OSS account.**
key	**The key of the OSS account.**
bucket	**The bucket where the data file is located. You must use OSS to create the bucket before data import.**

Parameter	Description
prefix	The prefix of the path name corresponding to the data file. Prefixes are directly matched and cannot be controlled by regular expressions. The prefix, filepath, and dir parameters are mutually exclusive and only one parameter can be specified at a time. • All OSS files containing the specified prefix will be imported if you create a READABLE external table for data import. - The following files will be imported if you set prefix to test/filename: ■ test/filename ■ test/filenamexxx ■ test/filename/aa ■ test/filenameyyy/aa ■ test/filenameyyy/bb/aa - Only the following file out of the preceding files will be imported if you set prefix to test/filename/: test/filename/aa • The exported files are uniquely named based on this parameter if you create a WRITABLE external table for data export. Note: One or more files can be exported for each data node. The names of exported files are in the `prefix_tablename_uuid.x` format. uuid indicates a timestamp in microseconds as an int64 value. x indicates the node ID. You can use an external table for multiple export operations. Each export operation is assigned a uuid value. The files exported during each operation share a uuid value.

Parameter	Description
dir	The virtual folder path in OSS. The prefix, filepath, and dir parameters are mutually exclusive and only one parameter can be specified at a time. · A folder path must end with a forward slash (/) such as `test/mydir/`. · If you use this parameter when creating an external table for data import, all files under the specified virtual directory will be imported , excluding its subdirectories and files in those subdirectories. Unlike filepath, dir does not require you to specify the names of files in the directory. · All data will be exported to multiple files in the specified directory if this parameter is used in creating an external table for data export. The names of exported files are in the `filename.x` format, where x is a number. The values of x may be inconsecutive.

Parameter	Description

filepath	The file name that contains a path in OSS. The prefix , filepath, and dir parameters are mutually exclusive and only one parameter can be specified at a time. You can only specify the filepath parameter when you create a READABLE external table for data import.
	· The file name includes the file path, but does not include the bucket name.
	· The file name specified for data import must be in the `filename` or `filename.x` format. The values of x must be consecutive numbers starting from 1.
	For example, if you set filename to filepath and OSS contains the following files, the imported files include filename, filename.1, and filename.2. Because filename.3 does not exist, filename.4 will not be imported.
	`filename`
	`filename.1`
	`filename.2`
	`filename.4`

Table 12-5: Import mode parameters

Parameter	Description
async	Specifies whether to asynchronously load data. · Asynchronous data import is enabled by default. You can set async to false or f to disable asynchronous data import. · Enables the worker thread to load data from OSS to accelerate the import performance. The default import mode is asynchronous mode. · Asynchronous data import consumes more hardware resources than normal data import.

Parameter	Description
compressiontype	The compression format of the imported file. Valid values: · None: specifies whether to import files that are not compressed. This is the default value. · gzip: specifies whether to import files of the GZIP format. Only the GZIP format is supported.
compressionlevel	The compression level of the files written to OSS. The value range is 1 to 9 and the default value is 6.

Table 12-6: Export mode parameters

Parameter	Description
oss_flush_block_size	The buffer size for the data written to OSS at a time. The default value is 32 MB, and the value range is 1 MB to 128 MB.
oss_file_max_size	The maximum size of the file written to OSS. If the limit is exceeded, subsequent data is written in another file. The default value is 1024 MB, and the value range is 8 MB to 4000 MB.
num_parallel_worker	The number of parallel compression threads for the data written to OSS. The value range is 1 to 8 and the default value is 3.

In addition, you must pay attention to the following items for the export mode:

- **WRITABLE is the keyword of the external table for data export. You must specify this keyword when creating an external table.**
- **Only the prefix and dir parameters are supported for data export. The filepath parameter is not supported.**
- **You can use the DISTRIBUTED BY clause to write data from segments to OSS based on the specified distribution keys.**

Other common parameters

The following error-tolerance parameters can be used for data import and export:

Table 12-7: Error-tolerance parameters

Parameter	Description
oss_connect_timeout	The connection timeout period. Unit: seconds. Default value: 10.
oss_dns_cache_timeout	The DNS timeout period. Unit: seconds. Default value : 60.
oss_speed_limit	The minimum tolerable rate. Default value: 1024 bit/s (1 Kbit/s).
oss_speed_time	The maximum tolerable time. Unit: seconds. Default value: 15.

If the default values are used for the preceding parameters, a timeout occurs when the transmission rate is lower than 1 Kbit/s for 15 consecutive seconds. For more information, *see* **Troubleshooting in OSS SDK reference.**

The other parameters are compatible with the original external table syntax of Greenplum Database. For more information about the syntax, see *Greenplum Database official documentation on external table syntax*. **These parameters include:**

- **FORMAT: indicates the supported file format, such as TEXT and CSV.**
- **ENCODING: indicates the data encoding format of a file, such as UTF-8.**
- **LOG ERRORS: indicates that the clause can ignore imported erroneous data and write the data to error_table. You can also use the count parameter to specify the error reporting threshold.**

Examples

```
#Create a READABLE external table of OSS create readable external table
 ossexample (date text, time text, open float, high float, low float,
volume int) location('oss://oss-cn-hangzhou.aliyuncs.com prefix=osstest
/example id=XXX key=XXX bucket=testbucket compressiontype=gzip') FORMAT
 'csv' (QUOTE '''' DELIMITER E'\t') ENCODING 'utf8' LOG ERRORS INTO
my_error_rows SEGMENT REJECT LIMIT 5;create readable external table
ossexample (date text, time text, open float, high float, low float,
volume int) location('oss://oss-cn-hangzhou.aliyuncs.com dir=osstest
/ id=XXX key=XXX bucket=testbucket') FORMAT 'csv' LOG ERRORS SEGMENT
 REJECT LIMIT 5;create readable external table ossexample (date text
, time text, open float, high float, low float, volume int) location
('oss://oss-cn-hangzhou.aliyuncs.com filepath=osstest/example.csv id=
XXX key=XXX bucket=testbucket') FORMAT 'csv' LOG ERRORS SEGMENT REJECT
 LIMIT 5;#Create a WRITABLE external table of OSS create WRITABLE
external table ossexample_exp (date text, time text, open float, high
float, low float, volume int) location('oss://oss-cn-hangzhou.aliyuncs
.com prefix=osstest/exp/outfromhdb id=XXX key=XXX bucket=testbucket
') FORMAT 'csv' DISTRIBUTED BY (date);create WRITABLE external table
 ossexample_exp (date text, time text, open float, high float, low
```

```
float, volume int) location('oss://oss-cn-hangzhou.aliyuncs.com dir=
osstest/exp/ id=XXX key=XXX bucket=testbucket') FORMAT 'csv' DISTRIBUTE
D BY (date);#Create a heap table to load data create table example
 (date text, time text, open float, high float, low float, volume
int) DISTRIBUTED BY (date);#Load data from ossexample to example in
parallel insert into example select * from ossexample;#Export data from
 example to OSS insert into ossexample_exp select * from example;#Each
 segment is involved. #Each segment pulls data from OSS in parallel.
The redistribution motion node calculates the hash value of the data
 and distributes the hash value to the corresponding segment. That
segment then imports the data to the database through the insert node
. explain insert into example select * from ossexample; QUERY PLAN
 ------------------------------------------------------------------------
Insert (slice0; segments: 4) (rows=250000 width=92) -> Redistribute
 Motion 4:4 (slice1; segments: 4) (cost=0.00.. 11000.00 rows=250000
 width=92) Hash Key: ossexample.date -> External Scan on ossexample
 (cost=0.00.. 11000.00 rows=250000 width=92)(4 rows)# The segment
imports the local data to OSS. Data redistribution is not performed.
explain insert into ossexample_exp select * from example; QUERY PLAN
 ------------------------------------------------------------------ Insert
 (slice0; segments: 3) (rows=1 width=92) -> Seq Scan on example (cost=0.
00.. 0.00 rows=1 width=92)(2 rows)
```

TEXT and CSV format description

The following parameters specify the formats of files read from and written to OSS. You can specify the parameters in the external DDL parameters.

- **The string \n is used as a line delimiter or line break for TEXT and CVS files.**
- **DELIMITER: specifies the delimiter of columns.**
  - **If the DELIMITER parameter is specified, the QUOTE parameter must also be specified.**
  - **Recommended column delimiters include commas (,), vertical bars (|), \t, and other special characters.**
- **QUOTE: encloses user data that contains special characters by column.**
  - **Strings that contain special characters will be enclosed by QUOTE to differenti ate user data and the control characters.**
  - **To optimize the efficiency, it is unnecessary to enclose data such as integers in QUOTE characters.**
  - **QUOTE cannot be the same string as specified in DELIMITER. The default value of QUOTE is double quotation marks (").**
  - **User data that contains QUOTE characters must also contain ESCAPE characters to differentiate the user data from code for the machine.**

- **ESCAPE: specifies the escape character.**
  - **Place an escape character before a special character that needs to be escaped to indicate that it is not a special character.**
  - **If ESCAPE is not specified, the default value is the same as QUOTE.**
  - **You can also use other characters as ESCAPE characters such as backslashes (\), which is used by MySQL.**

Default control characters for TEXT and CSV files

Table 12-8: Default control characters for TEXT and CSV files

Control character	TEXT	CSV
DELIMITER	\t (tab)	, (comma)
QUOTE	" (double quotation mark)	" (double quotation mark)
ESCAPE	N/A	Same as QUOTE
NULL	\N (backslash-N)	Empty string without quotation marks

All control characters must be single-byte characters.

SDK troubleshooting

If an error occurs during the import or export process, the error log contains the information as described in the following *Table 12-9: Error log information* **table.**

Table 12-9: Error log information

Keyword	Description
code	The HTTP status code of the error request.
error_code	The error code returned by OSS.
error_msg	The error message returned by OSS.
req_id	The UUID that identifies the request. If you require assistance in solving a problem, you can submit a ticket that contains the req_id of the failed request to OSS developers.

References

- *Greenplum Database official documentation on database external table syntax*

- *Greenplum Database official documentation on database table creation syntax*

12.3.5.2 Import data from MySQL

You can use the mysql2pgsql tool to migrate tables from MySQL to AnalyticDB for PostgreSQL, Greenplum Database, PostgreSQL, or PPAS.

Background information

mysql2pgsql connects the source MySQL database to the destination AnalyticDB for PostgreSQL database, queries the data to be exported from the MySQL database, and then imports the data to the destination database by using the \COPY command . The tool supports multi-thread importing (each worker thread imports a part of database tables).

To download the binary installation package of mysql2pgsql, click *here*.

To view instructions on source code compilation of mysql2pgsql, click *here*.

Procedure

1. **Modify the my.cfg configuration file to configure the connection information of source and destination databases.**
   a. **Modify the connection information of the source MySQL database.**

   Note:
   You must have the read permissions on all user tables.

   ```
   [src.mysql]
   host = "192.168.1.1"
   port = "3306"
   user = "test"
   password = "test"
   db = "test"
   encodingdir = "share"
   encoding = "utf8"
   ```

   b. **Modify the connection information of the destination PostgreSQL, PPAS, or AnalyticDB for PostgreSQL database.**

   Note:
   You must have the write permissions on the destination table.

   ```
   [desc.pgsql]
   ```

```
connect_string = "host=192.168.1.2 dbname=test port=3432  user=
test password=pgsql"
```

2. **Import data by using mysql2pgsql.**

```
./mysql2pgsql -l <tables_list_file> -d -n -j <number of threads> -s
 <schema of target able>
```

Table 12-10: Parameters

Parameter	Description

-l	**Optional. It is used to specify a text file that contains tables to be synchronized. If you do not specify this parameter, all the tables in the database that is specified in the configuration file will be synchronized.** `<tables_list_file>` **is the name of a file that contains a collection of tables that need to be synchronized and conditions for table queries. The content format is as follows:**
	`table1 : select * from table_big where column1 < '2016 -08-05'`
	`table2 :`
	`table3`
	`table4: select column1, column2 from tableX where column1 ! = 10`
	`table5: select * from table_big where column1 >= '2016 -08-05'`

-d	**Optional. It indicates the table creation DDL statement that only creates the destination table but does not synchronize data.**
-n	**Optional. It must be used along with -d to specify that the table partition definition is not included in the DDL statement.**
-j	**Optional. It is used to specify the number of threads used for data synchronization. If you do not specify this parameter, five threads will be used concurrently.**
-s	**Optional. It is used specify the schema of the destination table. Only one schema at a time can be specified by the command. If you do not specify the parameter, the data is imported into the table under the public schema.**

Typical usage

Full database migration

1. Obtain the DDL statements of the corresponding destination table by running the following command.

   ```
   ./mysql2pgsql -d
   ```

2. Create a table in the destination database based on these DDL statements with the distribution key information added.
3. Run the following command to synchronize all tables:

   ```
   ./mysql2pgsql
   ```

   This command will migrate the data from all MySQL tables in the database that is specified in the configuration file to the destination database. Five threads (default) are used during the process to read and import the data from all tables involved.

Partial table migration

1. Create a new file tab_list.txt and enter the following content:

   ```
   t1
   t2 : select * from t2 where c1 > 138888
   ```

2. Run the following command to synchronize the specified t1 and t2 tables (note that for the t2 table, only data that meets the c1 > 138888 condition is migrated):

   ```
   ./mysql2pgsql -l tab_list.txt
   ```

12.3.5.3 Import data from PostgreSQL

You can use the pgsql2pgsql tool to migrate tables across AnalyticDB for PostgreSQL, Greenplum Database, PostgreSQL, and PPAS.

Context

pgsql2pgsql supports the following features:

- Full migration across PostgreSQL, PPAS, Greenplum Database, and AnalyticDB for PostgreSQL.
- Full migration and incremental migration from PostgreSQL or PPAS (version 9.4 or later) to PostgreSQL or PPAS.

You can download the software packages from the *dbsync project* library.

- To download the binary installation package of pgsql2pgsql, click *here*.
- To view instructions on source code compilation of pgsql2pgsql, click *here*.

Procedure

1. **Modify the my.cfg configuration file to configure the connection information of source and destination databases.**

   a) **Modify the connection information of the source PostgreSQL database.**

   > **Note:**
   > **In the connection information of the source PostgreSQL database, we recommend that you set the user to the owner of the source database.**

   ```
   [src.pgsql]
   connect_string = "host=192.168.1.1 dbname=test port=3432  user=
   test password=pgsql"
   ```

   b) **Modify the connection information of the local temporary PostgreSQL database.**

   ```
   [local.pgsql]
   connect_string = "host=192.168.1.2 dbname=test port=3432  user=
   test2 password=pgsql"
   ```

   c) **Modify the connection information of the destination PostgreSQL database.**

   > **Note:**
   > **You must have the write permissions on the destination table.**

   ```
   [desc.pgsql]
   connect_string = "host=192.168.1.2 dbname=test port=3432  user=
   test3 password=pgsql"
   ```

   > **Note:**
   > - **If you need to synchronize incremental data, the source database must have the permissions to create replication slots.**
   > - **PostgreSQL 9.4 and later support logic flow replication, and therefore source databases of the versions support incremental migration. The kernel supports logic flow replication only if you configure the following kernel parameters.**
   >
   > **wal_level = logical**
   >
   > **max_wal_senders = 6**
   >
   > **max_replication_slots = 6**

2. **Use pgsql2pgsql to perform full database migration.**

   ```
   ./pgsql2pgsql
   ```

   The migration program will migrate the table data of all the users from the source PostgreSQL database to the destination PostgreSQL database by default.

3. **View the status information.**

   You can view the status information in a single migration process by connecting to the local temporary database. The information is stored in the db_sync_status table, including the start and end time of the full migration, the start time of the incremental migration, and the status of incremental synchronization.

12.3.5.4 Import data by using the \COPY command

You can use the `\COPY` command to import the data of local text files into AnalyticDB for PostgreSQL databases. The local text files must be formatted, such as files that use commas (,), colons (.), or special characters as delimiters.

Context

- **Parallel writing of massive data is not available because the `\COPY` command performs serial data writing by using the master node. If you need to import a large amount of data in parallel, you can use the OSS-based data import method.**
- **The `\COPY` command is a psql instruction. If you use the database statement `COPY` instead of the `\COPY` command, you must note that only stdin is supported. This COPY statement does not support file because the root user does not have the superuser permissions to perform operations on files.**
- **AnalyticDB for PostgreSQL also allows you to use JDBC to execute the COPY statement. The CopyIn method is encapsulated in JDBC. For more information, see *Interface CopyIn*.**
- **For more information about the usage of the COPY statement, see *COPY*.**

Procedure

Import data by using the following sample code:

```
\COPY table [(column [, ...])] FROM {'file' | STDIN}
     [ [WITH]
       [OIDS]
       [HEADER]
       [DELIMITER [ AS ] 'delimiter']
       [NULL [ AS ] 'null string']
       [ESCAPE [ AS ] 'escape' | 'OFF']
       [NEWLINE [ AS ] 'LF' | 'CR' | 'CRLF']
       [CSV [QUOTE [ AS ] 'quote']
```

```
            [FORCE NOT NULL column [, ...]]
       [FILL MISSING FIELDS]
       [[LOG ERRORS [INTO error_table] [KEEP]
       SEGMENT REJECT LIMIT count [ROWS | PERCENT] ]
\COPY {table [(column [, ...])] | (query)} TO {'file' | STDOUT}
      [ [WITH]
        [OIDS]
        [HEADER]
        [DELIMITER [ AS ] 'delimiter']
        [NULL [ AS ] 'null string']
        [ESCAPE [ AS ] 'escape' | 'OFF']
        [CSV [QUOTE [ AS ] 'quote']
             [FORCE QUOTE column [, ...]] ]
      [IGNORE EXTERNAL PARTITIONS ]
```

12.4 Databases

12.4.1 Overview

The operations based on the Greenplum Database in AnalyticDB for PostgreSQL are the same as those in the Greenplum Database, including schema, supported data types, and user permissions. Apart from some operations exclusive to the Greenplum Database such as the partition keys and AO tables, you can refer to PostgreSQL for other operations.

References

- *Pivotal Greenplum Official Documentation*
- *Greenplum 4.3 Best Practices*

12.4.2 Create a database

After you log on to the AnalyticDB for PostgreSQL instance, you can execute SQL statements to create databases.

Like in PostgreSQL, you can execute SQL statements to create databases in AnalyticDB for PostgreSQL. For example, after psql is connected to Greenplum, execute the following statements:

```
=> create database mygpdb;
CREATE DATABASE
=> \c mygpdb
psql (9.4.4, server 8.3devel)
You are now connected to database "mygpdb" as user "mygpdb".
```

12.4.3 Create a partition key

AnalyticDB for PostgreSQL is a distributed database and data is distributed across all the data nodes. You must create partition keys to distribute the data. The

partition keys are vital to query performance. Partition keys are used to ensure even data distribution. Proper selection of the keys will help significantly improve query performance.

Specify a partition key

In AnalyticDB for PostgreSQL, tables are distributed across all segments. The distribution rules are hash or random. You must specify the partition key when creating a table. Imported data will be distributed to the specific segment based on the hash value calculated by the partition key.

```
=> create table vtbl(id serial, key integer, value text, shape cuboid,
 location geometry, comment text) distributed by (key);
CREATE TABLE
```

If you do not specify the partition key (that means a statement without the `distributed by (key)` field), AnalyticDB for PostgreSQL randomly allocates the ID field by using the round-robin algorithm.

Rules for selecting the partition key

- **Select evenly distributed columns or multiple columns to prevent data skew.**
- **Select fields commonly used for connection operations, especially for highly concurrent statements.**
- **Select the condition columns that feature high concurrency queries and high filterability.**
- **Do not use random distributions.**

12.4.4 Construct data

In some test scenarios, you need to construct data to fill the database.

1. **Create a function that generates random strings.**

```
CREATE OR REPLACE FUNCTION random_string(integer) RETURNS text AS $
body$
SELECT array_to_string(array
                       (SELECT substring('0123456789ABCDEFGHIJ
KLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz'
                                         FROM (ceil(random()*62))::
int
                                         FOR 1)
                        FROM generate_series(1, $1)), '');
$body$
```

```
LANGUAGE SQL VOLATILE;
```

2. **Create a partition key.**

```
CREATE TABLE tbl(id serial, KEY integer, locate geometry, COMMENT
text) distributed by (key);
```

3. **Construct data.**

```
INSERT INTO tbl(KEY, COMMENT, locate)
 SELECT
     KEY,
     COMMENT,
     ST_GeomFromText(locate) AS locate
 FROM
     (SELECT
         (a + 1) AS KEY,
         random_string(ceil(random() * 24)::integer) AS COMMENT,
         'POINT(' || ceil(random() * 36 + 99) || ' ' || ceil(random
() * 24 + 50) || ')' AS locate
     FROM
         generate_series(0, 99999) AS a)
     AS t;
```

12.4.5 Query data

This topic describes the query statements and how to view the query plans.

Query statement sample

```
  => select * from tbl where key = 751;
  | id  | key | value | shape        | locate
           | comment        |
  +-----+-----+-------+-------------
+-----------------------------------------+----------------+
  | 751 | 751 | red   | 0101000000000000000C05B400000000000004A40 |
 B9hPhjeNWPqV |
  (1 row)
  Time: 513.101 ms
```

View a query plan

```
  => explain select * from tbl where key = 751;
  Gather Motion 1:1  (slice1; segments: 1)  (cost=0.00.. 1519.28
 rows=1 width=53)
    ->  Seq Scan on tbl  (cost=0.00.. 1519.28 rows=1 width=53)
          Filter: key = 751
  Settings:  effective_cache_size=8GB; gp_statistics_use_fkeys=on
```

```
  Optimizer status: legacy query optimizer
```

12.4.6 Manage extensions

You can use extensions to expand database features. AnalyticDB for PostgreSQL enables you to manage extensions.

Extension types

AnalyticDB for PostgreSQL supports the following extensions:

- **PostGIS: supports geographic information data.**
- **MADlib: supports the machine learning function library.**
- **fuzzystrmatch: supports the fuzzy matching of strings.**
- **orafunc: compatible with some Oracle functions.**
- **oss_ext: supports reading data from OSS.**
- **hll: collects statistics by using the HyperLogLog algorithm.**
- **pljava: supports compiling user-defined functions (UDF) in the PL/Java language.**
- **pgcrypto: supports cryptographic hash functions.**
- **intarray: supports integer array-related functions, operators, and indexes.**

Create an extension

Execute the following statements to create an extension:

```
CREATE EXTENSION <extension name>;
CREATE SCHEMA <schema name>;
CREATE EXTENSION IF NOT EXISTS <extension name> WITH SCHEMA <schema
name>;
```

Note:

Before you create the MADlib extension, you must create the plpythonu extension first.

```
CREATE EXTENSION plpythonu;
CREATE EXTENSION madlib;
```

Delete an extension

Execute the following statements to delete an extension:

```
DROP EXTENSION <extension name>;
```

```
DROP EXTENSION IF EXISTS <extension name> CASCADE;
```

Note:
If there are objects dependent on the extension, you need to add the CASCADE keyword to delete all dependent objects.

12.4.7 Manage users and permissions

This topic describes how to manage users and permissions in AnalyticDB for PostgreSQL.

Manage users

The system will prompt you to specify an initial username and password during instance creation. This initial user is the root user. After the instance is created, you can use the root user account to connect to the database. The system also creates superusers such as aurora and replicator for internal management.

You can run the `\du+` command to view the information of all the users after you connect to the database by using the client tool of PostgreSQL or Greenplum. Example:

```
postgres=> \du+
                                   List of roles
  Role name   |            Attributes             | Member of |
 Description
--------------+-----------------------------------+-----------
+---------------
 root_user    |                                   |      | rds_superuser
 ...
```

AnalyticDB for PostgreSQL does not provide superuser permissions, but offers a similar role of RDS_SUPERUSER, which is consistent with the permission system in ApsaraDB RDS for PostgreSQL. Therefore, the root user (such as the root_user in the preceding example) has the RDS_SUPERUSER permissions. You can only identify this permission attribute by viewing the user description.

The root user has the following permissions:

- **Creates databases and users and performs actions such as LOGIN, but excluding the SUPERUSER permissions.**
- **Views and modifies the data tables of users other than the superuser and performs actions such as SELECT, UPDATE, DELETE, and changing owner.**

- **Views the connection information of users other than the superuser, cancels their SQL statements, and kills their connections.**
- **Executes the CREATE EXTENSION and DROP EXTENSION statements to create and delete extensions.**
- **Creates other users that have the RDS_SUPERUSER permissions. Example:**

```
CRATE ROLE root_user2 RDS_SUPERUSER LOGIN PASSWORD 'xyz' ;
```

Manage permissions

You can manage permissions at the database, schema, and table levels. For example , if you want to grant the read permissions on tables to a user and revoke the write permissions, you can execute the following statements:

```
GRANT SELECT ON TABLE t1 TO normal_user1;
REVOKE UPDATE ON TABLE t1 FROM normal_user1;
REVOKE DELETE ON TABLE t1 FROM normal_user1;
```

12.4.8 Manage JSON data

JavaScript Object Notation (JSON) has become a basic data type in the Internet and IoT fields. For more information about JSON, visit *JSON official website*. PostgreSQL support for JSON has been well developed. Optimized by Alibaba Cloud, AnalyticDB for PostgreSQL supports the JSON type based on the PostgreSQL syntax.

Check whether the current version supports JSON

Execute the following statement to check whether the current version supports JSON:

```
=> SELECT '""'::json;
```

If the following output is displayed, it indicates the JSON type is supported and the instance is ready for use. If the operation fails, restart the instance.

```
 json
------
 ""
(1 row)
```

If the following output is displayed, it indicates the JSON type is not supported.

```
ERROR:  type "json" does not exist
LINE 1: SELECT '""'::json;
```

```
                  ^
```

The preceding command converts data from the string type to the JSON type. PostgreSQL supports operations on JSON data based on this conversion.

JSON conversion in the database

Database operations include reading and writing. The written data is typically converted from the string type to the JSON type. The contents of a string must meet the JSON standard, such as strings, digits, arrays, and objects. Example:

String

```
=> SELECT '"hijson"'::json;
 json
-------
 "hijson"
(1 row)
```

`::` is used for explicit type conversion in PostgreSQL, Greenplum, and AnalyticDB for PostgreSQL. The database calls the input function in JSON type during the conversion. Therefore, the JSON format check is performed as follows:

```
=> SELECT '{hijson:1024}'::json;
ERROR:  invalid input syntax for type json
LINE 1: SELECT '{hijson:1024}'::json;
               ^
DETAIL:  Token "hijson" is invalid.
CONTEXT:  JSON data, line 1: {hijson...
=>
```

In the preceding example, `hijson` must be enclosed in double quotation marks (") because JSON requires the KEY value to be a string. A syntax error is returned when `{hijson:1024}` is entered.

Apart from explicit type conversion, database records can also be converted to JSON.

Typically, JSON is not used to just a string or a number, but an object that contains one or more key-value pairs. AnalyticDB for PostgreSQL can support most JSON scenarios after data is converted from the string type to objects. Example:

```
=> select row_to_json(row('{"a":"a"}', 'b'));
           row_to_json
--------------------------------
 {"f1":"{\"a\":\"a\"}","f2":"b"}
(1 row)
=> select row_to_json(row('{"a":"a"}'::json, 'b'));
        row_to_json
---------------------------
 {"f1":{"a":"a"},"f2":"b"}
```

```
(1 row)
```

You can see the differences between the string and JSON here. The whole record is conveniently converted into the JSON type.

JSON data types

- **Object**

  The object is the most frequently used data type in JSON. Example:

  ```
  => select '{"key":"value"}'::json;
        json
  -----------------
   {"key":"value"}
  (1 row)
  ```

- **Integer and floating point number**

  JSON only supports three types of numeric values: integer, floating point number , and constant expression. AnalyticDB for PostgreSQL supports all the three types .

  ```
  => SELECT '1024'::json;
   json
  ------
   1024
  (1 row)
  => SELECT '0.1'::json;
   json
  ------
   0.1
  (1 row)
  ```

  The following information is required in some special situations:

  ```
  => SELECT '1e100'::json;
   json
  -------
   1e100
  (1 row)
  => SELECT '{"f":1e100}'::json;
      json
  -------------
   {"f":1e100}
  (1 row)
  ```

  Extra-long numbers are also supported. Example:

  ```
  => SELECT '9223372036854775808'::json;
          json
  ---------------------
   9223372036854775808
  ```

```
  (1 row)
```

- **Array**

```
=> SELECT '[[1,2], [3,4,5]]'::json;
        json
  ------------------
   [[1,2], [3,4,5]]
  (1 row)
```

Operators

Operators supported by JSON

```
=> select oprname,oprcode from pg_operator where oprleft = 3114;
 oprname |          oprcode
---------+---------------------------
 ->      | json_object_field
 ->>     | json_object_field_text
 ->      | json_array_element
 ->>     | json_array_element_text
 #>      | json_extract_path_op
 #>>     | json_extract_path_text_op
(6 rows)
```

Basic usage

```
=> SELECT '{"f":"1e100"}'::json -> 'f';
 ? column?
----------
 "1e100"
(1 row)
=> SELECT '{"f":"1e100"}'::json ->> 'f';
 ? column?
----------
 1e100
(1 row)
=> select '{"f2":{"f3":1},"f4":{"f5":99,"f6":"stringy"}}'::json#>array
['f4','f6'];
 ? column?
-----------
 "stringy"
(1 row)
=> select '{"f2":{"f3":1},"f4":{"f5":99,"f6":"stringy"}}'::json#>'{f4,
f6}';
 ? column?
-----------
 "stringy"
(1 row)
=> select '{"f2":["f3",1],"f4":{"f5":99,"f6":"stringy"}}'::json#>>'{f2
,0}';
 ? column?
----------
 f3
```

```
(1 row)
```

JSON functions

Supported JSON function

```
postgres=# \df *json*
                                                                List of
functions
   Schema   |           Name            | Result data type |
         Argument data types                    |  Type
------------+---------------------------+------------------
+-----------------------------------------------------------+--------
 pg_catalog | array_to_json             | json             | anyarray
                                                 | normal
 pg_catalog | array_to_json             | json             | anyarray
, boolean                                                  | normal
 pg_catalog | json_array_element        | json             | from_json
 json, element_index integer                             | normal
 pg_catalog | json_array_element_text   | text             | from_json
 json, element_index integer                             | normal
 pg_catalog | json_array_elements       | SETOF json       | from_json
 json, OUT value json                                    | normal
 pg_catalog | json_array_length         | integer          | json
                                                 | normal
 pg_catalog | json_each                 | SETOF record     | from_json
 json, OUT key text, OUT value json                      | normal
 pg_catalog | json_each_text            | SETOF record     | from_json
 json, OUT key text, OUT value text                      | normal
 pg_catalog | json_extract_path         | json             | from_json
 json, VARIADIC path_elems text[]                        | normal
 pg_catalog | json_extract_path_op      | json             | from_json
 json, path_elems text[]                                 | normal
 pg_catalog | json_extract_path_text    | text             | from_json
 json, VARIADIC path_elems text[]                        | normal
 pg_catalog | json_extract_path_text_op | text             | from_json
 json, path_elems text[]                                 | normal
 pg_catalog | json_in                   | json             | cstring
                                                 | normal
 pg_catalog | json_object_field         | json             | from_json
 json, field_name text                                   | normal
 pg_catalog | json_object_field_text    | text             | from_json
 json, field_name text                                   | normal
 pg_catalog | json_object_keys          | SETOF text       | json
                                                 | normal
 pg_catalog | json_out                  | cstring          | json
                                                 | normal
 pg_catalog | json_populate_record      | anyelement       | base
anyelement, from_json json, use_json_as_text boolean | normal
 pg_catalog | json_populate_recordset   | SETOF anyelement | base
anyelement, from_json json, use_json_as_text boolean | normal
 pg_catalog | json_recv                 | json             | internal
                                                 | normal
 pg_catalog | json_send                 | bytea            | json
                                                 | normal
 pg_catalog | row_to_json               | json             | record
                                                 | normal
 pg_catalog | row_to_json               | json             | record,
boolean                                                    | normal
 pg_catalog | to_json                   | json             |
anyelement                                                          | normal
```

```
(24 rows)
```

Basic usage

```
=> SELECT array_to_json('{{1,5},{99,100}}'::int[]);
  array_to_json
-----------------
 [[1,5],[99,100]]
(1 row)
=> SELECT row_to_json(row(1,'foo'));
     row_to_json
---------------------
 {"f1":1,"f2":"foo"}
(1 row)
=> SELECT json_array_length('[1,2,3,{"f1":1,"f2":[5,6]},4]');
 json_array_length
-------------------
                 5
(1 row)
=> select * from json_each('{"f1":[1,2,3],"f2":{"f3":1},"f4":null,"f5
":99,"f6":"stringy"}') q;
 key |   value
-----+-----------
 f1  | [1,2,3]
 f2  | {"f3":1}
 f4  | null
 f5  | 99
 f6  | "stringy"
(5 rows)
=> select json_each_text('{"f1":[1,2,3],"f2":{"f3":1},"f4":null,"f5":"
null"}');
  json_each_text
------------------
 (f1,"[1,2,3]")
 (f2,"{""f3"":1}")
 (f4,)
 (f5,null)
(4 rows)
=> select json_array_elements('[1,true,[1,[2,3]],null,{"f1":1,"f2":[7,
8,9]},false]');
  json_array_elements
-----------------------
 1
 true
 [1,[2,3]]
 null
 {"f1":1,"f2":[7,8,9]}
 false
(6 rows)
create type jpop as (a text, b int, c timestamp);
=> select * from json_populate_record(null::jpop,'{"a":"blurfl","x":43
.2}', false) q;
   a    | b | c
--------+---+---
 blurfl |   |
(1 row)
=>     select * from json_populate_recordset(null::jpop,'[{"a":"blurfl
","x":43.2},{"b":3,"c":"2012-01-20 10:42:53"}]',false) q;
   a    | b |            c
--------+---+--------------------------
 blurfl |   |
        | 3 | Fri Jan 20 10:42:53 2012
```

```
(2 rows)
```

Code examples

Create a table

```
create table tj(id serial, ary int[], obj json, num integer);
=> insert into tj(ary, obj, num) values('{1,5}'::int[], '{"obj":1}', 5
);
INSERT 0 1
=> select row_to_json(q) from (select id, ary, obj, num from tj) as q;
                row_to_json
-------------------------------------------
 {"f1":1,"f2":[1,5],"f3":{"obj":1},"f4":5}
(1 row)
=> insert into tj(ary, obj, num) values('{2,5}'::int[], '{"obj":2}', 5
);
INSERT 0 1
=> select row_to_json(q) from (select id, ary, obj, num from tj) as q;
                row_to_json
-------------------------------------------
 {"f1":1,"f2":[1,5],"f3":{"obj":1},"f4":5}
 {"f1":2,"f2":[2,5],"f3":{"obj":2},"f4":5}
(2 rows)
```

Join multiple tables

```
create table tj2(id serial, ary int[], obj json, num integer);
=> insert into tj2(ary, obj, num) values('{2,5}'::int[], '{"obj":2}',
5);
INSERT 0 1
=> select * from tj, tj2 where tj.obj->>'obj' = tj2.obj->>'obj';
 id |  ary  |    obj    | num | id |  ary  |    obj    | num
----+-------+-----------+-----+----+-------+-----------+-----
  2 | {2,5} | {"obj":2} |   5 |  1 | {2,5} | {"obj":2} |   5
(1 row)
=> select * from tj, tj2 where json_object_field_text(tj.obj, 'obj')
 = json_object_field_text(tj2.obj, 'obj');
 id |  ary  |    obj    | num | id |  ary  |    obj    | num
----+-------+-----------+-----+----+-------+-----------+-----
  2 | {2,5} | {"obj":2} |   5 |  1 | {2,5} | {"obj":2} |   5
(1 row)
```

Use JSON function index

```
CREATE TEMP TABLE test_json (
       json_type text,
       obj json
);
=> insert into test_json values('aa', '{"f2":{"f3":1},"f4":{"f5":99,"
f6":"foo"}}');
INSERT 0 1
=> insert into test_json values('cc', '{"f7":{"f3":1},"f8":{"f5":99,"
f6":"foo"}}');
INSERT 0 1
=> select obj->'f2' from test_json where json_type = 'aa';
 ? column?
----------
 {"f3":1}
(1 row)
```

```
=> create index i on test_json (json_extract_path_text(obj, '{f4}'));
CREATE INDEX
=> select * from test_json where json_extract_path_text(obj, '{f4}') =
 '{"f5":99,"f6":"foo"}';
 json_type |                 obj
-----------+--------------------------------------
 aa        | {"f2":{"f3":1},"f4":{"f5":99,"f6":"foo"}}
(1 row)
```

Note:

JSON data cannot be used as the partition key and does not support JSON aggregate functions.

Example of using Python to access the database:

```
#! /bin/env python
import time
import json
import psycopg2
def gpquery(sql):
    conn = None
    try:
        conn = psycopg2.connect("dbname=sanity1x2")
        conn.autocommit = True
        cur = conn.cursor()
        cur.execute(sql)
        return cur.fetchall()
    except Exception as e:
        if conn:
            try:
                conn.close()
            except:
                pass
            time.sleep(10)
        print e
    return None
def main():
    sql = "select obj from tj;"
    #rows = Connection(host, port, user, pwd, dbname).query(sql)
    rows = gpquery(sql)
    for row in rows:
        print json.loads(row[0])
if __name__ == "__main__":
    main()
```

12.4.9 Use HyperLogLog

AnalyticDB for PostgreSQL is highly optimized by Alibaba Cloud, and not only has the features of Greenplum Database, but also supports HyperLogLog. It is suitable for industries with requirements similar to Internet advertising and estimation analysis, which require quick estimation of business metrics such as PV and UV.

Create a HyperLogLog extension

You can execute the following statement to create a HyperLogLog extension:

```
CREATE EXTENSION hll;
```

Basic types

- **Execute the following statement to create a table containing the hll field:**

  ```
  create table agg (id int primary key,userids hll);
  ```

- **Execute the following statement to convert int to hll_hashval:**

  ```
  select 1::hll_hashval;
  ```

Basic operators

- **The hll type supports** =, !=, <>, ||, **and** #**.**

  ```
  select hll_add_agg(1::hll_hashval) = hll_add_agg(2::hll_hashval);
  select hll_add_agg(1::hll_hashval) || hll_add_agg(2::hll_hashval);
  select #hll_add_agg(1::hll_hashval);
  ```

- **The hll_hashval type supports** =, !=, **and** <>**.**

  ```
  select 1::hll_hashval = 2::hll_hashval;
  select 1::hll_hashval <> 2::hll_hashval;
  ```

Basic functions

- **Hash functions such as Hll_hash_boolean, hll_hash_smallint, and hll_hash_b igint.**

  ```
  select hll_hash_boolean(true);
  select hll_hash_integer(1);
  ```

- **hll_add_agg: converts the int format to the hll format.**

  ```
  select hll_add_agg(1::hll_hashval);
  ```

- **hll_union: aggregates the hil fields.**

  ```
  select hll_union(hll_add_agg(1::hll_hashval),hll_add_agg(2::hll_hashva
  l));
  ```

- **hll_set_defaults: sets the precision.**

  ```
  select hll_set_defaults(15,5,-1,1);
  ```

- **hll_print: displays debug information.**

  ```
  select hll_print(hll_add_agg(1::hll_hashval));
  ```

Example

```
create table access_date (acc_date date unique, userids hll);
```

```
insert into access_date select current_date, hll_add_agg(hll_hash_i
nteger(user_id)) from generate_series(1,10000) t(user_id);
insert into access_date select current_date-1, hll_add_agg(hll_hash_i
nteger(user_id)) from generate_series(5000,20000) t(user_id);
insert into access_date select current_date-2, hll_add_agg(hll_hash_i
nteger(user_id)) from generate_series(9000,40000) t(user_id);
postgres=# select #userids from access_date where acc_date=current_da
te;
     ? column?
------------------
 9725.85273370708
(1 row)
postgres=# select #userids from access_date where acc_date=current_da
te-1;
     ? column?
------------------
 14968.6596883279
(1 row)
postgres=# select #userids from access_date where acc_date=current_da
te-2;
     ? column?
------------------
 29361.5209149911
(1 row)
```

12.4.10 Use the CREATE LIBRARY statement

AnalyticDB for PostgreSQL introduces the CREATE LIBRARY and DROP LIBRARY statements to allow you to import custom software packages.

Syntax

```
CREATE LIBRARY library_name LANGUAGE [JAVA] FROM oss_location OWNER
ownername
CREATE LIBRARY library_name LANGUAGE [JAVA] VALUES file_content_hex
OWNER ownername
DROP LIBRARY library_name
```

Table 12-11: Parameters

Parameter	Description
library_name	The name of the library to be installed. If the name of the library to be installed has the same name as an existing library, you must delete the existing library before installing the new one.
LANGUAGE [JAVA]	The programming language to be used. Only PL/Java is supported.

Parameter	Description
oss_location	The location of the package. You can specify the OSS bucket and object names. Only one object can be specified and the specified object cannot be a compressed file. The format is as follows: `oss://oss_endpoint filepath=[folder/[folder/]...]/file_name id=userossid key=userosskey bucket=ossbucket`
file_content_hex	The content of the file. The byte stream is in hexadecimal notation. For example, 73656c6563742031 indicates the hexadecimal byte stream of " select 1". You can use this syntax to import packages without using OSS.
ownername	Specifies the user.
DROP LIBRARY	Deletes a library.

Examples

- Example 1: Install a JAR package named analytics.jar.

```
create library example language java from 'oss://oss-cn-hangzhou.aliyuncs.com filepath=analytics.jar id=xxx key=yyy bucket=zzz';
```

- Example 2: Import the file content with the byte stream in hexadecimal notation.

```
create library pglib LANGUAGE java VALUES '73656c6563742031' OWNER "myuser";
```

- Example 3: Delete a library.

```
drop library example;
```

- Example 4: View installed libraries.

```
select name, lanname from pg_library;
```

12.4.11 Create and use the PL/Java UDF

AnalyticDB for PostgreSQL supports compiling and uploading JAR software packages written in PL/Java languages, and using these JAR packages to create user-defined functions (UDFs). The PL/Java language supported by AnalyticDB for PostgreSQL is Community Edition PL/Java 1.5.0 and the JVM version is 1.8. This topic describes how to create a PL/Java UDF. For more PL/Java examples, see *PL/Java code*. For the compiling method, see *PL/Java documentation*.

Procedure

1. **In AnalyticDB for PostgreSQL, execute the following statement to create a PL/Java extension. You only need to execute the statement once for each database.**

```
create extension pljava;
```

2. **Compile the UDF based on your business needs. For example, you can use the following code to compile the Test.java file:**

```
public class Test
 {
     public static String substring(String text, int beginIndex,
                int endIndex)
             {
                     try {
                         Process process = null;
                         process = Runtime.getRuntime().exec("echo
Test running");
                     } catch (Exception e) {
                        return "" + e;
                     }
                     return text.substring(beginIndex, endIndex);
             }
 }
```

3. **Compile the manifest.txt file.**

```
 Manifest-Version: 1.0
 Main-Class: Test
 Specification-Title: "Test"
 Specification-Version: "1.0"
 Created-By: 1.7.0_99
 Build-Date: 01/20/2016 21:00 AM
```

4. **Run the following commands to compile and package the program.**

```
javac Test.java
 jar cfm analytics.jar manifest.txt Test.class
```

5. **Upload the analytics.jar file generated in step 4 to OSS by using the following OSS console command.**

```
osscmd put analytics.jar oss://zzz
```

6. **In AnalyticDB for PostgreSQL, execute the CREATE LIBRARY statement to import the file to AnalyticDB for PostgreSQL.**

```
create library example language java from 'oss://oss-cn-hangzhou.
aliyuncs.com filepath=analytics.jar id=xxx key=yyy bucket=zzz';
```

Note:

You can only use the filepath variable in the CREATE LIBRARY statement to import files. You can import one file at a time. In addition, the CREATE LIBRARY statement also supports byte streams to import files without using OSS. For more information, see *Use the CREATE LIBRARY statement*.

7. **In AnalyticDB for PostgreSQL, execute the following statements to create and use the UDF.**

```
create table temp (a varchar) distributed randomly;
 insert into temp values ('my string');
 create or replace function java_substring(varchar, int, int)
returns varchar as 'Test.substring' language java;
 select java_substring(a, 1, 5) from temp;
```

12.5 Tables

12.5.1 Create a table

You can create tables in the database.

Syntax

The complete syntax for creating a table is as follows. Not all clauses are required. Use the clauses that suits your business needs.

```
CREATE [[GLOBAL | LOCAL] {TEMPORARY | TEMP}] TABLE table_name (
[ { column_namedata_type [ DEFAULT default_expr ]
   [column_constraint [ ... ]
[ ENCODING ( storage_directive [,...] ) ]
]
   | table_constraint
   | LIKE other_table [{INCLUDING | EXCLUDING}
                      {DEFAULTS | CONSTRAINTS}] ...}
   [, ... ] ]
   )
   [ INHERITS ( parent_table [, ... ] ) ]
   [ WITH ( storage_parameter=value [, ... ] )
   [ ON COMMIT {PRESERVE ROWS | DELETE ROWS | DROP} ]
   [ TABLESPACE tablespace ]
   [ DISTRIBUTED BY (column, [ ... ] ) | DISTRIBUTED RANDOMLY ]
   [ PARTITION BY partition_type (column)
       [ SUBPARTITION BY partition_type (column) ]
          [ SUBPARTITION TEMPLATE ( template_spec ) ]
       [...]
    ( partition_spec )
        | [ SUBPARTITION BY partition_type (column) ]
          [...]
    ( partition_spec
      [ ( subpartition_spec
           [(...)]
       ) ]
```

```
    )
```

The `column_constraint` **clause can be defined as follows:**

```
[CONSTRAINT constraint_name]
   NOT NULL | NULL
   | UNIQUE [USING INDEX TABLESPACE tablespace]
            [WITH ( FILLFACTOR = value )]
   | PRIMARY KEY [USING INDEX TABLESPACE tablespace]
                 [WITH ( FILLFACTOR = value )]
   | CHECK ( expression )
   | REFERENCES table_name [ ( column_name [, ... ] ) ]
            [ key_match_type ]
            [ key_action ]
```

The `storage_directive` **clause of columns can be defined as follows:**

```
   COMPRESSTYPE={ZLIB | QUICKLZ | RLE_TYPE | NONE}
    [COMPRESSLEVEL={0-9} ]
    [BLOCKSIZE={8192-2097152} ]
```

The `storage_parameter` **clause of tables can be defined as follows:**

```
   APPENDONLY={TRUE|FALSE}
   BLOCKSIZE={8192-2097152}
   ORIENTATION={COLUMN|ROW}
   CHECKSUM={TRUE|FALSE}
   COMPRESSTYPE={ZLIB|QUICKLZ|RLE_TYPE|NONE}
   COMPRESSLEVEL={0-9}
   FILLFACTOR={10-100}
   OIDS[=TRUE|FALSE]
```

The `table_constraint` **clause can be defined as follows:**

```
   [CONSTRAINT constraint_name]
   UNIQUE ( column_name [, ... ] )
          [USING INDEX TABLESPACE tablespace]
          [WITH ( FILLFACTOR=value )]
   | PRIMARY KEY ( column_name [, ... ] )
                 [USING INDEX TABLESPACE tablespace]
                 [WITH ( FILLFACTOR=value )]
   | CHECK ( expression )
   | FOREIGN KEY ( column_name [, ... ] )
            REFERENCES table_name [ ( column_name [, ... ] ) ]
            [ key_match_type ]
            [ key_action ]
            [ key_checking_mode ]
```

Valid values of `key_match_type`**:**

```
    MATCH FULL
  | SIMPLE
```

Valid values of `key_action`**:**

```
    ON DELETE
  | ON UPDATE
```

```
   | NO ACTION
   | RESTRICT
   | CASCADE
   | SET NULL
   | SET DEFAULT
```

Valid values of `key_checking_mode`**:**

```
     DEFERRABLE
   | NOT DEFERRABLE
   | INITIALLY DEFERRED
   | INITIALLY IMMEDIATE
```

Valid values of `partition_type`**:**

```
     LIST
   | RANGE
```

The `partition_specification` **clause can be defined as follows:**

```
partition_element [, ...]
```

The `partition_element` **clause can be defined as follows:**

```
    DEFAULT PARTITION name
   | [PARTITION name] VALUES (list_value [,...] )
   | [PARTITION name]
      START ([datatype] 'start_value') [INCLUSIVE | EXCLUSIVE]
      [ END ([datatype] 'end_value') [INCLUSIVE | EXCLUSIVE] ]
      [ EVERY ([datatype] [number | INTERVAL] 'interval_value') ]
   | [PARTITION name]
      END ([datatype] 'end_value') [INCLUSIVE | EXCLUSIVE]
      [ EVERY ([datatype] [number | INTERVAL] 'interval_value') ]
[ WITH ( partition_storage_parameter=value [, ... ] ) ]
[ TABLESPACE tablespace ]
```

The `subpartition_spec` **or** `template_spec` **clause can be defined as follows:**

```
subpartition_element [, ...]
```

The `subpartition_element` **clause can be defined as follows:**

```
    DEFAULT SUBPARTITION name
   | [SUBPARTITION name] VALUES (list_value [,...] )
   | [SUBPARTITION name]
      START ([datatype] 'start_value') [INCLUSIVE | EXCLUSIVE]
      [ END ([datatype] 'end_value') [INCLUSIVE | EXCLUSIVE] ]
      [ EVERY ([datatype] [number | INTERVAL] 'interval_value') ]
   | [SUBPARTITION name]
      END ([datatype] 'end_value') [INCLUSIVE | EXCLUSIVE]
      [ EVERY ([datatype] [number | INTERVAL] 'interval_value') ]
[ WITH ( partition_storage_parameter=value [, ... ] ) ]
```

```
[ TABLESPACE tablespace ]
```

The `storage_parameter` clause of partitions can be defined as follows:

```
    APPENDONLY={TRUE|FALSE}
    BLOCKSIZE={8192-2097152}
    ORIENTATION={COLUMN|ROW}
    CHECKSUM={TRUE|FALSE}
    COMPRESSTYPE={ZLIB|QUICKLZ|RLE_TYPE|NONE}
    COMPRESSLEVEL={1-9}
    FILLFACTOR={10-100}
    OIDS[=TRUE|FALSE]
```

Parameters

The *Table 12-12: Table creation parameters* table describes the key parameters for creating a table.

Table 12-12: Table creation parameters

Parameter	Description
TABLE_NAME	**The name of the table to be created.**
column_name	**The name of a column to be created in the new table.**
data_type	**The data type of the column.** **For columns that contain textual data, set the data type to `VARCHAR` or `TEXT`. We do not recommend the `CHAR` type.**
DEFAULT default_expr	**Specifies a default value for the column. The system assigns default values for all columns that do not have values. The values can be any variable-free expression. Subqueries or cross-references to other columns in the table are not allowed. The data type of the default expression must match the data type of the column. If a column does not have a default value, the default value is null.**

Parameter	Description
ENCODING storage_directive	Specifies the type of compression and block size for the column data. This clause is valid only for append-optimized, column-oriented tables. Column compression settings are inherited from the table level to the partition level to the subpartition level. The lowest-level settings have priority over the inherited settings .
INHERITS	Specifies that all columns in the new table automatically inherit a parent table. You can use `INHERITS` to create a persistent relationship between the new child table and its parent table. Schema modifications to the parent table are typically applied to the child table. When the parent table is scanned, the data of the child table is scanned as well.
LIKE other_table	Specifies a table from which the new table automatically copies all column names, data types, NOT NULL constraints, and distribution policies. Storage properties such as append-optimized or partition structure are not copied. Unlike `INHERITS`, the new table is completely decoupled from the original table after the new table is created.
CONSTRAINT constraint_name	Configures a column or table constraint. If a constraint is violated, the constraint name is displayed in the error messages, so constraint names can be used to communicate helpful information to client applications. Constraint names that contain spaces must be enclosed in double quotation marks (").
WITH (storage_option=value)	Configures storage options for the table or its indexes.

Parameter	Description
ON COMMIT	The operation that the system performs on the temporary tables at the end of a transaction. Valid values: · PRESERVE ROWS: No special action is taken. After the transaction is completed, the data is retained. The data is released only when the session is disconnected. · DELETE ROWS: All rows in the temporary table are deleted. · DROP: The temporary table is deleted.
TABLESPACE tablespace	Specifies the name of the tablespace in which the new table is to be created. If not specified, the default tablespace of the database is used.
DISTRIBUTED BY	Specifies the distribution policy for the database. · `DISTRIBUTED BY (column, [ ... ] )`: specifies the partition key. The system uses hash distribution based on the distribution key. To evenly distribute the data, you must specify the partition key to the primary key of the table or a unique column or a set of columns. · `DISTRIBUTED RANDOMLY`: distributes data randomly. **Note:** We recommend that you do not to use random distribution.
PARTITION BY	Configures the partition key to partition the table. Partitioning large tables improves data access efficiency. To partition a table is to create a top-level (parent) table and multiple lower-level (child) tables. A parent table is always empty once a partition table is created. The data is stored in the lowest-level child tables. In a multi-level partition table, data is only stored in the lowest-level sub-partitions. Valid values: RANGE, LIST, and a combination of the two.
SUBPARTITION BY	Configures a multi-level partition table.

Parameter	Description
SUBPARTITION TEMPLATE	**You can specify a sub-partition template to create sub-partitions (lower-level child tables). This sub-partition template is applied to all parent partitions to ensure the same sub-partition structure.**

Examples

Create a table and configure the partition key. The primary key is the default partition key in AnalyticDB for PostgreSQL.

```
CREATE TABLE films (
code        char(5) CONSTRAINT firstkey PRIMARY KEY,
title       varchar(40) NOT NULL,
did         integer NOT NULL,
date_prod   date,
kind        varchar(10),
len         interval hour to minute
);

CREATE TABLE distributors (
did    integer PRIMARY KEY DEFAULT nextval('serial'),
name   varchar(40) NOT NULL CHECK (name <> '')
);
```

Create a compressed table and configure the partition key.

```
CREATE TABLE sales (txn_id int, qty int, date date)
WITH (appendonly=true, compresslevel=5)
DISTRIBUTED BY (txn_id);
```

Create a three-level partition table by using sub-partition templates of each level and the default partition.

```
CREATE TABLE sales (id int, year int, month int, day int,
region text)
DISTRIBUTED BY (id)
PARTITION BY RANGE (year)

  SUBPARTITION BY RANGE (month)
    SUBPARTITION TEMPLATE (
       START (1) END (13) EVERY (1),
       DEFAULT SUBPARTITION other_months )

  SUBPARTITION BY LIST (region)
    SUBPARTITION TEMPLATE (
       SUBPARTITION usa VALUES ('usa'),
       SUBPARTITION europe VALUES ('europe'),
       SUBPARTITION asia VALUES ('asia'),
       DEFAULT SUBPARTITION other_regions)

( START (2008) END (2016) EVERY (1),
```

```
  DEFAULT PARTITION outlying_years);
```

12.5.2 Principles and scenarios of row store, column store, heap tables, and AO tables

AnalyticDB for PostgreSQL supports row store, column store, heap tables, and AO tables. This topic describes their principles and scenarios.

Row store and column store

Table 12-13: Comparison

Dimension	Row store	Column store
Concept	Row store stores data in the form of rows. Each row is a tuple. To read a column, you must deform all the columns before this column. Therefore, the costs for accessing the first and the last columns are different.	Column store stores data in the form of columns. Each column corresponds to one file or a batch of files. The cost of reading any column is the same, but if you need to read multiple columns, you must access multiple files. The more columns you access, the higher the overheads are.
Compression ratio	Low.	High.
Cost of reading any column	Columns with larger column numbers cost more.	Same.
Vector computing and JIT architecture	Not suitable. Not suitable for batch computation.	Suitable. More efficient when accessing and obtaining statistics of a batch of data.

Dimension	Row store	Column store
Scenario	If you need to perform a large number of update and delete operations because of the OLTP requirements, for example, to query table details, and multiple columns are returned, you can use row store. If you have diversified requirements, you can use partition tables. For example, if you partition the data based on time, you can use row store to query the details of recent data and use column store to obtain more historical data statistics.	If you need data statistics because of the OLAP requirements, you can use column store. If you need a higher compression ratio, you can use column store.

Heap tables

A heap table is heap storage. All changes to the heap table generate redo logs that can be used to restore data by time point. However, heap tables cannot implement logical incremental backup, because any data block in the table may be changed and it is not convenient to record the position by using the heap storage.

When a transaction is finished, commit logs and redo logs are used to ensure that it is reliable. You can also build mirror nodes by using redo logs to achieve data redundancy.

Append-optimized (AO) tables

AO tables are used to append data for storage. When you delete the updated data, you can use another bitmap file to mark the row to be deleted, and use the bit and offset to determine whether a row is deleted.

When the transaction is finished, you must call the fsync function to record the offset of the data block that performs the last write operation. Even if the data block contains only one record, a new data block will be appended for the next transactio n. The data block is synchronized to the mirror node for data redundancy.

Therefore, AO tables are not suitable for small transactions because the fsync function is called at the end of each transaction, and this data block will not be reused even if there is space left.

AO tables are suitable for OLAP scenarios, batch data writing, high compression ratio, and logical backup that supports incremental backup. During backup, you only need to record the offset from the backup and the bitmap deletion mark for each full backup, which requires little space.

Usage scenarios of heap tables

- **When multiple small transactions are handled, use a heap table.**
- **When you need to restore the data by time point, use a heap table.**

Usage scenarios of AO tables

- **When you need to use column store, use an AO table.**
- **When data is written in batches, use an AO table.**

12.5.3 Enable the column store and compression features

If you want to improve the performance, speed up imports, or reduce the cost of tables with infrequent updates and multiple fields, we recommend that you use column store and compression. This will increase the compression ratio threefold when guaranteeing performance, and the import speed is usually faster.

To enable the column store and compression features, you must specify the column store and compression options when creating the table. For example, you can add the following clause to the CREATE statement to enable the two features. For more information about the table creation syntax, see *Create a table*.

```
with (APPENDONLY=true, ORIENTATION=column, COMPRESSTYPE=zlib,
COMPRESSLEVEL=5, BLOCKSIZE=1048576, OIDS=false)
```

Note:

AnalyticDB for PostgreSQL only supports zlib and RLE_TYPE compression algorithms. If you specify the quicklz algorithm, it is automatically converted to zlib.

12.5.4 Add a field to a column store table and set the default value

This topic describes how to add a field to a column store table and set the default value for the field, and how to use the ANALYZE statement to view the impact of updated data on the size of the column store table.

Context

In a column store table, each column is stored as a file, and two columns in the same row correspond to each other by using the offset. For example, if you add two fields of the INT8 type, you can quickly locate column B from column A by using the offset.

When you add the field, AO tables are not rewritten. If an AO table contains the records of deleted data, the added field must be filled with the deleted records before using the relative offset.

Procedure

1. **Create three AO column store tables.**

```
postgres=# create table tbl1 (id int, info text) with (appendonly=
true, blocksize=8192, compresstype=none, orientation=column);
NOTICE:  Table doesn't have 'DISTRIBUTED BY' clause -- Using column
named 'id' as the Greenplum Database data distribution key for this
table.
HINT:  The 'DISTRIBUTED BY' clause determines the distribution of
data. Make sure column(s) chosen are the optimal data distribution
key to minimize skew.
CREATE TABLE

postgres=# create table tbl2 (id int, info text) with (appendonly=
true, blocksize=8192, compresstype=none, orientation=column);
NOTICE:  Table doesn't have 'DISTRIBUTED BY' clause -- Using column
named 'id' as the Greenplum Database data distribution key for this
table.
HINT:  The 'DISTRIBUTED BY' clause determines the distribution of
data. Make sure column(s) chosen are the optimal data distribution
key to minimize skew.
CREATE TABLE

postgres=# create table tbl3 (id int, info text) with (appendonly=
true, blocksize=8192, compresstype=none, orientation=column);
NOTICE:  Table doesn't have 'DISTRIBUTED BY' clause -- Using column
named 'id' as the Greenplum Database data distribution key for this
table.
```

```
HINT:  The 'DISTRIBUTED BY' clause determines the distribution of
data. Make sure column(s) chosen are the optimal data distribution
key to minimize skew.
CREATE TABLE
```

2. **Insert 10 million records to the first two tables and 20 million records to the third one.**

```
postgres=# insertinto tbl1 select generate_series(1,10000000),'test
';
INSERT 0 10000000
postgres=# insert into tbl2 select generate_series(1,10000000),'test
';
INSERT 0 10000000
postgres=# insert into tbl3 select generate_series(1,20000000),'test
';
INSERT 0 20000000
```

3. **Analyze the tables and display their sizes.**

```
postgres=# analyze tbl1;
ANALYZE
postgres=# analyze tbl2;
ANALYZE
postgres=# analyze tbl3;
ANALYZE

postgres=# select pg_size_pretty(pg_relation_size('tbl1'));
 pg_size_pretty
----------------
 88 MB
(1 row)
postgres=# select pg_size_pretty(pg_relation_size('tbl2'));
 pg_size_pretty
----------------
 88 MB
(1 row)
postgres=# select pg_size_pretty(pg_relation_size('tbl3'));
 pg_size_pretty
----------------
 173 MB
(1 row)
```

4. **Update all the data in the first table. Display the size after the update. The size is twice as large as the size before the update.**

```
postgres=# update tbl1 set info='test';
UPDATE 10000000
postgres=# analyze tbl1;
ANALYZE
postgres=# select pg_size_pretty(pg_relation_size('tbl1'));
 pg_size_pretty
----------------
 173 MB
(1 row)
```

5. **Add fields to the three tables and set the default values.**

```
postgres=# alter table tbl1 add column c1 int8 default 1;
```

```
ALTER TABLE
postgres=# alter table tbl2 add column c1 int8 default 1;
ALTER TABLE
postgres=# alter table tbl3 add column c1 int8 default 1;
ALTER TABLE
```

6. **Analyze the tables and view the table sizes.**

```
postgres=# analyze tbl1;
ANALYZE
postgres=# analyze tbl2;
ANALYZE
postgres=# analyze tbl3;
ANALYZE

postgres=# select pg_size_pretty(pg_relation_size('tbl1'));
 pg_size_pretty
----------------
 325 MB
(1 row)

postgres=# select pg_size_pretty(pg_relation_size('tbl2'));
 pg_size_pretty
----------------
 163 MB
(1 row)

postgres=# select pg_size_pretty(pg_relation_size('tbl3'));
 pg_size_pretty
----------------
 325 MB
(1 row)
```

When you add fields to the AO tables, the number of records in the existing files prevails. Even if all the records are deleted, you must initialize the original data on the newly added fields.

12.5.5 Configure the table partition

For fact tables or large-sized tables in the database, we recommend that you configure table partitions.

Configure the table partition

You can use the table partitioning feature to delete data by using the `alter table drop partition` statement to delete all the data in a partition, and import data by using the `alter table exchange partition` statement to add a new data partition on a regular basis.

AnalyticDB for PostgreSQL supports range partitioning, list partitioning, and composite partitioning. Range partitioning only supports partitioning by fields of the numeric and date and time data types.

The following example shows a table that uses range partitioning.

```
CREATE TABLE LINEITEM (
L_ORDERKEY              BIGINT NOT NULL,
L_PARTKEY               BIGINT NOT NULL,
L_SUPPKEY               BIGINT NOT NULL,
L_LINENUMBER    INTEGER,
L_QUANTITY              FLOAT8,
L_EXTENDEDPRICE FLOAT8,
L_DISCOUNT              FLOAT8,
L_TAX                   FLOAT8,
L_RETURNFLAG    CHAR(1),
L_LINESTATUS    CHAR(1),
L_SHIPDATE              DATE,
L_COMMITDATE    DATE,
L_RECEIPTDATE   DATE,
L_SHIPINSTRUCT  CHAR(25),
L_SHIPMODE              CHAR(10),
L_COMMENT               VARCHAR(44)
)  WITH  (APPENDONLY=true, ORIENTATION=column, COMPRESSTYPE=zlib
, COMPRESSLEVEL=5, BLOCKSIZE=1048576, OIDS=false) DISTRIBUTED BY (
l_orderkey)
PARTITION BY RANGE (L_SHIPDATE) (START (date '1992-01-01') INCLUSIVE
END (date '2000-01-01') EXCLUSIVE EVERY (INTERVAL '1 month' ));
```

Principles of table partitioning

The purpose of partitioning is to minimize the amount of data to be scanned by queries, so the partitions must be associated with the query conditions.

- **Principle 1: Select the fields related to the query conditions to configure partitions. This reduces the amount of data to be scanned.**
- **Principle 2: When multiple query conditions exist, configure sub-partitions to further reduce the amount of data to be scanned.**

12.5.6 Configure the sort key

A sort key is an attribute of a table. Data on disks are stored in the order of the sort key.

Context

Sort keys have two major advantages:

- **Speed up and optimizes column-store operations. The min and max meta information the system collects seldom overlaps with each other, featuring good filterability.**
- **Eliminate the need to perform ORDER BY and GROUP BY operations. The data directly read from the disk is ordered as required by the sorting conditions.**

Create a table

```
Command:     CREATE TABLE
Description: define a new table
Syntax:
CREATE [[GLOBAL | LOCAL] {TEMPORARY | TEMP}] TABLE table_name (
[ { column_name data_type [ DEFAULT default_expr ]     [column_con
straint [ ... ]
[ ENCODING ( storage_directive [,...] ) ]
]
   | table_constraint
   | LIKE other_table [{INCLUDING | EXCLUDING}
                      {DEFAULTS | CONSTRAINTS}] ...}
   [, ... ] ]
   [column_reference_storage_directive [, ] ]
   )
   [ INHERITS ( parent_table [, ... ] ) ]
   [ WITH ( storage_parameter=value [, ... ] )
   [ ON COMMIT {PRESERVE ROWS | DELETE ROWS | DROP} ]
   [ TABLESPACE tablespace ]
   [ DISTRIBUTED BY (column, [ ... ] ) | DISTRIBUTED RANDOMLY ]
   [ SORTKEY (column, [ ... ] )]
   [ PARTITION BY partition_type (column)
       [ SUBPARTITION BY partition_type (column) ]
          [ SUBPARTITION TEMPLATE ( template_spec ) ]
       [...]
    ( partition_spec )
        | [ SUBPARTITION BY partition_type (column) ]
          [...]
    ( partition_spec
      [ ( subpartition_spec
           [(...)]
        ) ]
```

```
)
```

Example:

```
create table test(date text, time text, open float, high float, low
 float, volume int) with(APPENDONLY=true,ORIENTATION=column) sortkey (
volume);
```

Sort the table

```
VACUUM SORT ONLY [tablename]
```

Modify the sort key

This statement only modifies the catalog and does not sort the data. You must execute the `vaccum sort only` statement to sort the table.

```
ALTER [[GLOBAL | LOCAL] {TEMPORARY | TEMP}] TABLE table_name SET
SORTKEY (column, [ ... ] )
```

Example:

```
alter table test set sortkey (high,low);
```

12.6 Best practices

12.6.1 Configure memory and load parameters

You need to configure memory and load parameters to improve database stability.

Background information

AnalyticDB for PostgreSQL is an MPP database with high computational and resource requirements. It consumes all the resources provided to it. This allows AnalyticDB for PostgreSQL to have processing speeds but it also makes it easy to exceed its limits.

If the CPU, network, or hard disk exceed their limits, the worse-case scenario is a hardware bottleneck. However, if the memory limit is exceeded, the database may crash.

How to prevent OOM

Out of memory (OOM) indicates that the system is unable to provide sufficient memory requested by a process. The following prompt appears when OOM errors occur:

```
Out of memory (seg27 host.example.com pid=47093) VM Protect failed to
allocate 4096 bytes, 0 MB available
```

Causes

Possible causes of the OOM error include:

- **The memory of the database node is insufficient.**
- **Kernel parameters related to the memory of the operating system are incorrectly configured.**
- **Data skew occurs, causing a segment to request a large amount of memory.**
- **Query skew occurs. For example, if the grouping fields of some aggregate or window functions are not distribution keys, the data needs to be redistributed. After the redistribution, the data skews in a certain segment, which results in the segment requesting a large amount of memory.**

Solutions

1. **Modify the queries to request less memory.**
2. **Use the resource queue provided by AnalyticDB for PostgreSQL to limit the number of concurrent queries. Reduce the number of queries executed within the cluster at the same time to reduce the overall memory requested by the system.**
3. **Reduce the number of segments deployed on a host. For example, deploy eight segments instead of 16 segments on a host with 128 GB of memory. This allows each segment to use twice the amount of memory compared with the latter.**
4. **Increase the memory of a host.**
5. **Set the gp_vmem_protect_limit parameter to limit the maximum VMEM that can be used by a single segment. The memory size of a single host and the number of segments deployed on the host determine the maximum memory size that a single segment can use on average.**
6. **For SQL statements that have unpredictable memory usage, you can set the statement_mem parameter in the session to limit the memory usage of a single SQL statement, so as to prevent a single SQL statement from consuming all available memory.**
7. **Set the statement_mem parameter at the database level to apply to all the sessions in the database.**

8. **Use the resource queue to limit the maximum memory usage of the resource group. Add database users to the resource group to limit the overall memory used by these users.**

Configure memory-related parameters

Properly configuring the operating system, database parameters, and resource queue can effectively reduce the probability of OOM.

When calculating the average memory usage of a single segment on a single host, you must consider both the primary segment and the mirror segment. When the cluster encounters a host failure, the system switches the primary segment to the corresponding mirror segment. At this time, the number of segments on the host is more than usual. Therefore, you must consider the resources occupied by mirror segments during failover.

The following tables describe how to configure parameters of the operating system kernel and database to avoid OOM.

The *Table 12-14: Operating system kernel parameters* **describes the parameter configuration of the operating system kernel.**

Table 12-14: Operating system kernel parameters

Parameter	Description
huge page	Do not configure the huge page parameter of the system . AnalyticDB for PostgreSQL does not support the latest version of PostgreSQL and therefore does not support the huge page feature. The huge page parameter locks a part of the allocated memory. Database nodes cannot use this part of memory.

Parameter	Description
vm.overcommit_memory	If you use the swap space, set this parameter to 2. If you do not use the swap space, set this parameter to 0. Valid values: · 0: The requested memory space cannot exceed the difference between the total memory and the resident set size (RSS). An error is returned only when the memory is exceeded. · 1: Most processes use the malloc function to apply for the memory, but do not use all the memory applied. When this parameter is set to 1, the memory requested by the malloc function is allocated under all circumstances unless the memory is insufficient. · 2: Swap is also considered when the system calculates the memory space that can be applied for. You can apply for a large amount of memory even if the swap is triggered.
overcommit_ratio	The larger the value, the more the memory that the processes can apply for, and the less the memory reserved for the operating system. When this parameter is set to 2, the memory address that can be applied for cannot exceed swap + memory × overcommit_ratio.

For the parameter configuration of the database, see *Database parameters*.

Table 12-15: Database parameters

Parameter	Description
gp_vmem_protect_limit	Specifies the maximum memory that can be applied for all the processes on each segment. If the value is too great, it may result in system OOM or more serious problems. If the value is too small, SQL statements may not be executed even if the system has enough memory.

Parameter	Description
runaway_detector_activation_percent	Default value: 90. This value is specified as a percentage. When the memory used by any segment exceeds runaway_detector_activation_percent × gp_vmem_protect_limit/100, the query is terminated to prevent OOM. The termination starts from the query that occupies the maximum memory until the memory reaches a value lower than runaway_detector_activation_percent × gp_vmem_protect_limit/100. You can use the gp_toolkit.session_level_memory_consumption view to observe the memory usage of each session and runaway information.

Parameter	Description
statement_mem	Specifies the maximum memory that can be applied for a single SQL statement. When the maximum memory is exceeded, spill files are created. Default value: 125. Unit: MB. We recommend that you set this parameter according to the following formula: `(gp_vmem_protect_limit × 0.9)/max_expected_concurrent_queries` **Note:** · You can specify the statement_mem parameter in a session. If the current concurrency is low, and a session needs to run a query that requires a large amount of memory, you must specify this parameter in the session. · Statement_mem is suitable for limiting memory usage in low concurrency scenarios. For high concurrency scenarios, if you use statement_mem to limit the memory, each query is allocated a very small amount of memory. As a result, the performance of a small number of queries with high memory requirements in high concurrency scenarios is affected. We recommend that you use the resource queue to limit the maximum memory usage in high concurrency scenarios.
gp_workfile_limit_files_per_query	Specifies the maximum number of spill files that can be created by each query. When the memory requested by the query exceeds the statement_mem limit, spill files (also known as work files) are created, which is similar to the swap of the operating system. When the spill files used exceed the limit, the query will be terminated. Default value: 0, which indicates that an unlimited number of spill files can be created.

Parameter	Description
gp_workfile_compress_algorithm	Specifies the compression algorithm for the spill file. Valid values: none and zlib. Specifies the compression algorithm. The values optimize storage space or I/O by sacrificing CPU. You can set this parameter when the disk is insufficient or the spill file meets a write bottleneck.

Examples to calculate the memory parameters

The environment is as follows:

- **Host configuration:**

```
Total RAM = 256GB
SWAP = 64GB
```

- **Four hosts, each deployed with eight primary segments and eight mirror segments.**

  When a host fails, the eight primary segments are distributed to the remaining three hosts. A single host can be deployed with extra three primary segments from the failed host at most. A single host can be deployed with 11 primary segments at most.

1. **Calculate the total memory allocated to AnalyticDB for PostgreSQL by the operating system.**

   Reserve 7.5 GB and 5% of memory for the operating system and calculate the available memory for all applications, and divide the available memory by the empirical coefficient of 1.7.

```
gp_vmem = ((SWAP + RAM) - (7.5GB + 0.05 * RAM))/1.7
        = ((64 + 256) - (7.5 + 0.05 * 256))/1.7
        = 176
```

2. **Use the empirical coefficient of 0.026 to calculate overcommit_ratio.**

```
vm.overcommit_ratio = (RAM - (0.026 * gp_vmem))/RAM
                    = (256 - (0.026 * 176))/256
                    = .982

Set vm.overcommit_ratio to 98.
```

3. **Calculate gp_vmem_protect_limit (the protection parameter of the maximum memory usage for each segment), and divide gp_vmem by maximum_acting_primary_segments (the number of primary segments to be run on each other host after one host fails).**

```
gp_vmem_protect_limit calculation
gp_vmem_protect_limit = gp_vmem/maximum_acting_primary_segments
                      = 176/11
                      = 16GB
                      = 16384MB
```

Configure the resource queue

You can use resource queues to limit the number of concurrent queries and the total memory usage. When a query is running, it is added to the corresponding queue, and the resources used are recorded in the queue. The resource limit of the queue works for all sessions in the queue.

The resource queue in AnalyticDB for PostgreSQL is similar to the cgroup in Linux.

The syntax to create a resource queue is as follows:

```
Command:     CREATE RESOURCE QUEUE
Description: create a new resource queue for workload management
Syntax:
CREATE RESOURCE QUEUE name WITH (queue_attribute=value [, ... ])
where queue_attribute is:
   ACTIVE_STATEMENTS=integer
        [ MAX_COST=float [COST_OVERCOMMIT={TRUE|FALSE}] ]
        [ MIN_COST=float ]
        [ PRIORITY={MIN|LOW|MEDIUM|HIGH|MAX} ]
        [ MEMORY_LIMIT='memory_units' ]
|  MAX_COST=float [ COST_OVERCOMMIT={TRUE|FALSE} ]
        [ ACTIVE_STATEMENTS=integer ]
        [ MIN_COST=float ]
        [ PRIORITY={MIN|LOW|MEDIUM|HIGH|MAX} ]
        [ MEMORY_LIMIT='memory_units' ]
```

The *Table 12-16: Resource queue creation parameters* describes the parameters for creating the resource queue.

Table 12-16: Resource queue creation parameters

Parameter	Description
ACTIVE_STATEMENTS	**The number of SQL statements that are allowed to run (in the active status) concurrently.** **The value -1 indicates an unlimited number of SQL statements can run concurrently.**

Parameter	Description

MEMORY_LIMIT ' memory_units kB, MB or GB'	Specifies the maximum memory usage allowed by all SQL statements in the resource queue. The value -1 indicates unlimited memory usage, but it is easy to trigger OOM errors because it is limited by the database or system parameters mentioned in the preceding sections.
	
	The memory usage of SQL statements is limited by resource queues and parameters.
	
	· When the gp_resqueue_memory_policy parameter is set to none, the limit is the same as that in the Greenplum databases earlier than version 4.1.
	· When the gp_resqueue_memory_policy parameter is set to auto and you have specified the statement_mem parameter for a session or at the database level, the allowed memory of a single query will exceed the MEMORY_LIMIT of the resource queue.
	
	Example:
	
	`=> SET statement_mem='2GB';`
	`=> SELECT * FROM my_big_table WHERE column=' value' ORDER BY id;`
	`=> RESET statement_mem;`
	
	· The system parameter max_statement_mem can limit the maximum memory usage at the segment level. The memory requested by a single query cannot exceed max_statement_mem.
	
	You can modify the statement_mem parameter at the session level, but do not modify the max_statem ent_mem parameter. We recommend that you specify max_statement_mem as follows:
	
	`(seghost_physical_memory) / (average_nu mber_concurrent_queries)`
	
	· When the gp_resqueue_memory_policy parameter is set to eager_free, it indicates that the query is divided into several stages and the database allocates the memory requested in the current stage. For example, if a query requests 1 GB of memory in total but it only needs 100 MB during each stage, the database allocates 100 MB of memory to the query. You can use eager_free to reduce the possibility of insufficient memory for the query.

Parameter	Description
MAX_COST float	The maximum cost of the queries that are allowed to execute concurrently by the resource group. The cost is the estimated total cost in the SQL execution plan. The value of the parameter is specified as a floating point number (such as 100.0) or an exponent (such as 1e+2). The value -1 indicates the cost is unlimited.
COST_OVERCOMMIT boolean	Specifies whether the limit of max_cost can be exceeded when the system is idle. The value TRUE indicates the limit can be exceeded.
MIN_COST float	When the resources requested exceed the limit, the queries are queued. However, when the cost of a query is lower than the min_cost, the query can run without queuing.
PRIORITY={MIN|LOW|MEDIUM|HIGH|MAX}	The priority of the current resource queue. When resources are insufficient, CPU resources are allocated to the resource queue with a higher priority. The SQL statements in the resource queue with a higher priority can obtain CPU resources first. We recommend that you allocate users that initiate queries with high real-time requirements to resource queues with high priorities. This parameter is similar to the CPU resource group in the Linux cgroup and the time slice policy of real-time and common tasks.

Example of modifying resource queue limits:

```
ALTER RESOURCE QUEUE myqueue WITH (MAX_COST=-1.0, MIN_COST= -1.0);
```

Example of putting the user in the resource queue:

```
ALTER ROLE sammy RESOURCE QUEUE poweruser;
```

The following table describes the parameters of resource queues.

Table 12-17: Resource queue parameters

Parameter	Description
gp_resqueue_memory_policy	The memory management policy of the resource queue.
gp_resqueue_priority	Specifies whether to enable query prioritization. Valid values: · On · Off If this parameter is disabled, existing priority settings are not evaluated.
gp_resqueue_priority_cpucores_per_segment	Specifies the number of CPU cores allocated to each segment. For example, if an 8-core host is configured with two primary segments, you can set the parameter to 4. If there is no other node on the primary node, set the parameter to 8. When the CPU is preempted, the SQL statements running in the resource group with higher priority are allocated with CPU resources first.
gp_resqueue_priority_sweeper_interval	The interval at which CPU usage is recalculated for all active statements. The share value is calculated when the SQL statement is executed. You can calculate the share value based on the priority and gp_resqueue_priority_cpucores_per_segment. The smaller the value, the more frequent the calculation, and the better the result brought by the priority settings, the larger the overhead.

Tips for configuring resource queues

- **We recommend that you create a resource queue for each user.**

  The default resource queue of AnalyticDB for PostgreSQL is pg_default. If no queue is created, all users are assigned to pg_default. This operation is not recommended. We recommend that you create a resource queue for each user. Typically, a database user corresponds to a business. Different database users may correspond to different businesses or users, such as business users, analysts, developers, and DBAs.

- **We do not recommend that you use superusers to execute queries.**

  The queries initiated by superusers are not limited by the resource queue, but only by the preceding parameters. Therefore, if you want to use resource queues to limit the use of resources, we do not recommend that you use superusers to execute queries.
- **ACTIVE_STATEMENTS indicates the SQL statements that can be executed concurrently in the resource queue. When the cost of a query is lower than the min_cost, the query can run without queuing.**
- **You can specify the MEMORY_LIMIT parameter to set the allowed maximum memory usage of all the SQL statements in a resource queue. The statement_mem parameter has higher priority that can break through the limit of resource queues.**

  Note:
  The memory of all resource queues cannot exceed gp_vmem_protect_limit.
- **You can distinguish businesses by configuring the priorities of resource queues.**

  For example, the report forms are of top priority, followed by common businesses and analysts. In this case, you can create three resource queues with the medium, high, and max priorities, respectively.
- **If the resources requested in different time periods vary, you can use the crontab command to adjust the limits of resource queues periodically based on usage patterns.**

  For example, the queue of analysts has a higher priority during the day, and the queue of forms has a higher priority at night. AnalyticDB for PostgreSQL does not support resource limits by time period. Therefore, you can only deploy tasks externally by using the ALTER RESOURCE QUEUE statement.
- **You can use the view provided by gp_toolkit to observe the resource usage of the resource queues.**

```
gp_toolkit.gp_resq_activity

gp_toolkit.gp_resq_activity_by_queue

gp_toolkit.gp_resq_priority_backend

gp_toolkit.gp_resq_priority_statement

gp_toolkit.gp_resq_role
```

```
gp_toolkit.gp_resqueue_status
```

13 Data Transmission Service (DTS)

13.1 What is DTS?

Data Transmission Service (DTS) is a data service provided by Alibaba Cloud that supports data exchange between relational databases, OLAP databases, and other data sources.

DTS supports data migration, real-time data subscription, and real-time data synchronization. DTS can be used in multiple business scenarios, including interruption-free data migration, geo-disaster recovery, cross-border data synchronization, and cache update policies, helping you build a secure, scalable, and highly available data architecture.

- **DTS aims to help you with complex data interactions so that you can focus on upper-layer service development.**
- **DTS supports the following data sources:**
  - **Relational databases: MySQL and Oracle**
  - **OLAP databases: MaxCompute**

13.2 Log on to the DTS console

This topic uses the Google Chrome browser as an example to describe how to log on to the DTS console.

Prerequisites

- **Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://`IP address or domain name of the Apsara Stack console`/manage.**
- **We recommend that you use the Chrome browser.**

Procedure

1. **Open your browser.**

2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click LOGIN to go to the Dashboard page.
5. In the top navigation bar, choose  > Database > Data Transmission Service.
6. Click Console in the upper-right corner of the page.
7. Select your department and click DTS.

 **Note:**

- To use DTS, you must first authorize the selected department. You can click Authorize Now, and check whether your department is displayed in the Authorized Departments section. If your department is not displayed, select your department and click OK to authorize your department.
- You must select a valid department so that you can view the DTS instances created for the department.

13.3 Data migration

13.3.1 Overview

Data migration allows you to quickly migrate data between multiple data sources. Typical scenarios include data migration to the cloud, data migration between instances within Alibaba Cloud, and database split and scale-out. This section

introduces the instance types and data source types supported by the data migration feature of DTS.

Instance types supported by data migration

Table *Data source types supported by data migration* **lists the instance types supported by the data migration feature.**

Data source types supported by data migration

Table *Data source types supported by data migration* **lists the data source types supported by the data migration feature.**

Table 13-1: Data source types supported by data migration

Source database	Schema migration	Full data migration	Incremental data migration
MySQL > ApsaraDB RDS for MySQL	Supported	Supported	Supported
Oracle > ApsaraDB RDS for MySQL	Supported	Supported	Supported

Data migration supports the following source instance types:

- **RDS instances**
- **Oracle instances**
- **User-created databases**

Data migration supports the following destination instance types:

- **RDS instances**

13.3.2 Create a data migration task

The data migration feature provided by DTS allows you to configure a migration task by using three steps. This topic describes how to configure a task to migrate data from a MySQL instance to an ApsaraDB RDS for MySQL instance. You can follow a similar procedure to configure a task to migrate data to or from databases with other storage engines.

Prerequisites

- You have created the destination database in the destination RDS instance.

  If the destination database does not exist in the destination RDS instance, DTS automatically creates a destination database during data migration. However, in either of the following two cases, you must manually create the destination database in the RDS console before configuring a migration task:

  - The database name does not meet the following requirements of the RDS instance: A database name must be 1 to 64 characters in length and can contain lowercase letters, digits, underscores (_), and hyphens (-). It must start with a letter and end with a letter or digit.
  - The destination database has a different name from the source database.
- You have created the migration accounts.

  When configuring a migration task, you must provide the migration accounts of the source and destination instances. For more information about the database permissions required by each storage engine, see the *DTS documentation* .

  If you have not created a migration account for your source MySQL instance, follow the instructions in *Grant syntax* to create a migration account that meets the permission requirements.

  If you have not created a migration account for your destination ApsaraDB RDS for MySQL instance, follow the instructions *in* the Account management section of the ApsaraDB for RDS User Guide. You need to create a migration account and grant this account the read/write permission on the source and destination databases.

After the destination database and migration account are created, you can configure a migration task. Perform the following steps to configure a migration task:

Procedure

1. *Log on to the DTS console*.
2. In the left-side navigation pane, click Data Migration. On the page that appears, click Create Migration Task in the upper-right corner.
3. In the Create DTS Instances dialog box that appears, select a region, enter the number of data migration instances to be created, and click Create.
4. In the message that appears, click OK.

5. **In the migration task list, find the migration task that you created and click Configure Migration Task.**
6. **Configure the source and destination database instances, and click Set Whitelist and Next in the lower-right corner.**

   In this step, you need to configure the migration task name and the source and destination database instances. *Table 13-2: Parameters for configuring the migration task* **describes related parameters.**

   Table 13-2: Parameters for configuring the migration task

Type	Parameter	Description
Common parameter	Task Name	DTS automatically generates a name for each task. We recommend that you replace the default name with an informative name for easy identification.
Source database information	Instance Type	Valid values: · RDS Instance · User-Created Database with Public IP Address
	Database Type	Valid values: MySQL and Oracle.
	Hostname/IP Address	The IP address of your on-premises MySQL instance.
	Port Number	The port on which the MySQL instance is listening.
	Database Account	The account of the source database instance.
	Database Password	The password of the source database instance.
Destination database information	Instance Type	Valid values: · RDS Instance · User-Created Database with Public IP Address · DRDS Instance · PetaData
	RDS Instance ID	The ID of the destination database instance.

Type	Parameter	Description
	Database Account	The account of the destination database instance.
	Database Password	The password of the destination database instance.

7. Select migration types and objects to be migrated.

   In this step, you need to select migration types and objects to be migrated. The parameters are described as follows:

   - Migration types

     Available migration types are schema migration, full data migration, and incremental data migration.

     To perform a full data migration, select both Schema Migration and Full Data Migration for Migration Type.

     To perform a zero downtime migration, select Schema Migration, Full Data Migration, and Incremental Data Migration.

   - Objects to be migrated

     Select the objects to be migrated. Click the right arrow to add the selected objects to the Selected section on the right. An object to be migrated can be a database, table, or column.

     After objects are migrated to the destination database, the object names remain the same as that in the source database by default. If the object you migrate has different names in the source and destination instances, you must use the object name mapping feature provided by DTS. For more information, see *Database, table, and column name mapping*.

8. **Perform a precheck.**

   A precheck is required before you can start the migration task. A migration task can only be started after it passes the precheck.

   If the migration task fails the precheck, click the information icon correspond ing to each failed check item to view details, fix the issue, and run the precheck again.

   Click the information icon corresponding to each failed check item to view the cause and solution.

   After troubleshooting, select the task from the task list and perform the precheck again.

9. **Start the migration task.**

   After the migration task passes the precheck, you can start the migration task and check the migration status and progress in the task list.

Result

Note:

This is the complete procedure for creating the data migration task. You can follow a similar procedure to configure a task for migrating data to or from other types of instances or databases with other storage engines.

13.3.3 Precheck items

13.3.3.1 Source database connectivity

This check item checks whether the DTS server can connect to the source database for migration. DTS creates a connection to the source database by using the JDBC protocol. If the connection fails, the check item fails.

The source database connectivity check may fail for the following reasons:

- **An incorrect account or password is provided when a migration task is created.**

  Diagnostics:

  On any network-ready server that can connect to the source database, use the account and password specified for creating the migration task to connect to the source database through client software. Check whether the connection

succeeds. If an error is reported for the connection and the error message contains Access deny, the account or password is incorrect.

Troubleshooting:

Modify the migration task in the DTS console. Correct the account and password . Then re-run the precheck.

- The migration account of the source database implements access control based on source IP addresses.

  Diagnostics:

  - On any network-ready server that can connect to the source database, use the account and password specified for creating the migration task to connect to the source database through client software. Check whether the connection succeeds. If the connection succeeds, the source database has access restrictions based on IP addresses. Only allowed servers can connect to it. The IP address of the DTS server is not included in the whitelist of the source database, so the DTS server cannot connect to the source database.
  - If the source database is a MySQL database, you can access the source database by using the MySQL client. Run the `select host from mysql.user where user='Migration account', password='Migration account password'` command. If the query result is not %, the IP address of the DTS server is not included in the whitelist of the source database, which results in the connection failure.

  Troubleshooting:

  - If the source database is a MySQL database, run the `grant all on . to 'Migration account'@'%' identified by 'Migration account password';` command in the source database to re-authorize the migration account. Replace the migration account and password in this command with the real ones. After the account is authorized, re-run the precheck.
- A firewall is configured on the source database server.

  Diagnostics: If the source database is installed on a Linux server, run `iptables -L` in the shell to check whether a firewall has been configured on the server. If the source database is installed on a Windows server, perform the following operations: Open the Control Panel, click System and Security. On the System and

Security window that appears, click Windows Firewall. Check whether a firewall has been configured on the server.

Troubleshooting:

Disable the firewall and perform the precheck again.

- **There is no connectivity between the DTS server and the source database.**

  If none of the preceding cases applies, the check item may fail because there is no connectivity between the DTS server and the source database. In this case, contact the DTS engineers on duty.

13.3.3.2 Check the destination database connectivity

This check item checks whether the DTS server can connect to the destination database for migration. DTS creates a connection to the destination database by using the JDBC protocol. If the connection fails, the check item fails.

The destination database connectivity precheck may fail for the following reasons:

- **An incorrect account or password is provided when a migration task is created.**

  Diagnostics:

  On any network-ready server that can connect to the destination database, use the account and password specified for creating the migration task to connect to the destination database through client software. Check whether the connection succeeds. If an error is reported for the connection and the error message contains Access deny, the account or password is incorrect.

  Troubleshooting:

  Modify the migration task in the DTS console, correct the account and password, and perform the precheck again.

- **There is no connectivity between the DTS server and destination database.**

  If you check that the password and account are correct, the check item may fail because there is no connectivity between the DTS server and the destination database. In this case, contact the DTS engineers on duty.

13.3.3.3 Binlog configurations in the source database

Check whether binlogging is enabled for the source database

This check item is run only when incremental data is to be migrated between MySQL instances. This check item checks whether binlogging is enabled for the source database. If this check item fails, binlogging is not enabled for the source database.

Troubleshooting: Set `log_bin=mysql_bin` in the configuration file of the source database to enable binlogging. Restart the source database and re-run the precheck.

Check whether the binlog format is ROW in the source database

This check item is run only when incremental data is to be migrated between MySQL instances. This check item checks whether the binlog format is ROW in the source database. If this check item fails, the binlog format is not ROW in the source database.

Troubleshooting: Run the `set global binlog_format=ROW` command in the source database. Then, re-run the precheck. We recommend that you restart the source MySQL database after the modification. Otherwise, connected sessions may continue to be written in non-ROW mode, resulting in data loss.

Check whether a specified binlog file has been deleted from the source database

This check item is run only when incremental data is to be migrated between MySQL instances. This check item checks whether a specified binlog file has been deleted from the source database. If this check item fails, the binlog file does not exist in the source database.

Troubleshooting: Run the `PURGE BINARY LOGS TO "The name of the binlog file ranking the first place among all binlog files that have not been deleted"` command in the source database. Then, re-run the precheck.

For specific purge file names, see the precheck troubleshooting.

Check whether the binlog_row_image value of the MySQL source database is FULL

This check item is run only when incremental data is to be migrated between MySQL instances. This check item checks whether the binlog_row_image of the

source database is FULL, or whether the full image is recorded. If this check item fails, the binlog file of the source database does not record the full image.

Troubleshooting: Run the `set global binlog_row_image=FULL` command in the source database. Then, re-run the precheck.

13.3.3.4 Referential integrity constraint

This check item checks whether all the parent-child tables with foreign key dependencies among the objects to be migrated have been migrated, to avoid damaging the integrity of foreign key constraints.

If this check item fails, the failure cause is that the "parent table name" parent table on which the "child table name" table to be migrated is dependent has not been migrated.

Troubleshooting:

- **Do not migrate the child tables involved in the failed referential integrity constraint check. Modify the migration task and delete these child tables from the list of objects to be migrated. Then re-run the precheck.**
- **Migrate the parent tables for the child tables involved in the failed referential integrity constraint check. To do so, modify the migration task and add these parent tables to the list of objects to be migrated. Then re-run the precheck.**
- **Delete the foreign key dependencies of the child tables involved in the failed referential integrity constraint check. Modify the source database and delete the foreign key dependencies of these child tables. Then, re-run the precheck.**

13.3.3.5 Existence of Federated tables

This check item is run only when incremental data is to be migrated between MySQL instances. This check item checks whether any storage engines not supported by incremental data migration exist in the source database. Currently, incremental data migration does not support the Federated and the MRG_MyISAM storage engines.

If this check item fails and the error message "The Federated engine is used for the following source tables:" is displayed, the storage engine of some tables in the source database is Federated.

If this check item fails and the error message "The MRG_MyISAM engine is used for the following source tables:" is displayed, the storage engine of some tables in the source database is MRG_MyISAM.

Troubleshooting:

Modify the migration task by deleting the tables with the Federated or MRG_MyISAM storage engine from the list of objects to be migrated. Then create a separate migration task to implement schema migration and full data migration for these tables.

13.3.3.6 Permissions

Check the permissions granted to the migration account of the source database

This check item checks whether the migration account of the source database has the required permissions for data migration. For the migration permissions required by each type of database, see the Data Migration chapter.

Check the permissions granted to the migration account of the destination database

This check item checks whether the migration account of the source database has the required permissions for data migration. For the migration permissions required by each type of database, see the data migration chapter.

13.3.3.7 Object name conflict

This check item checks for duplicate object names in the destination and source database. If this check item fails, an object in the destination RDS instance has the same name as an object to be migrated. This causes the migration to fail.

When this check item fails, an error message is displayed indicating that an object in the destination database has the same name as an object to be migrated from the source database.

Troubleshooting:

- **Use the database and table name mapping feature provided by DTS to migrate the object to be migrated to another object with a different name in the destination database.**
- **In the destination database, delete or rename the object that has the same name as the object to be migrated.**

- **Modify the migration task and delete that object to be migrated from the list of objects to be migrated. Do not migrate this object.**

13.3.3.8 Schema existence

This check item checks whether the database to be migrated exists in the destination RDS instance. If no, DTS creates one automatically. However, under the following circumstances, the automatic database creation fails, and this check item prompts a failure:

- **The database name contains characters other than lowercase letters, digits, underscores (_), and hyphens (-).**

  The cause of the precheck failure is that the name of the source database does not comply with the requirements of RDS.

  Troubleshooting: On the database management page of the RDS console, create a database that complies with the requirements of RDS and grant the migration account the read and write permissions on the new database. Use the database name mapping feature provided by DTS to map the source database to the new database. Then, perform the precheck again.
- **The character set of the database is not UTF8, GBK, Latin1, or UTF-8MB4.**

  The cause of the precheck failure is that the character set of the source database does not comply with the requirements of RDS.

  Troubleshooting: On the database management page of the RDS console, create a database that complies with the requirements of RDS and grant the migration account the read and write permissions on the new database. If the new database and the database to be migrated have different names, you can use the database name mapping feature of DTS to map the database to be migrated to the new database. Then re-run the precheck.
- **The migration account of the destination database has no read and write permissions on the database to be migrated.**

  The cause of the precheck failure is that you are not authorized to operate on the source database.

  Troubleshooting: On the database management page of the RDS console, click the Account Management tab. Grant the migration account the read and write permissions on the source database. Then, perform the precheck again.

13.3.3.9 Source database server_id

This check item is run only when incremental data is to be migrated between MySQL instances. This check item checks whether server-id of the source database is set to an integer greater than 1.

If this check item fails, run the `set global server_id='an integer greater than 1'` command in the source database. Then run the precheck again.

13.3.3.10 Source database version

This check item checks whether the version of the source database is supported by DTS. *Table 13-3: Source database types and versions* **lists the source database versions supported by DTS.**

Table 13-3: Source database types and versions

Source database type	Supported version
MySQL	5.0, 5.1, 5.5, and 5.6. Only 5.1, 5.5, and 5.6 are supported for incremental data migration.

When the version check fails, you can only upgrade or downgrade the source database to the versions supported by DTS. Then re-run the precheck.

13.3.4 Migrate data from a local MySQL instance to an ApsaraDB RDS for MySQL instance

You can use DTS to migrate data from a local database to an ApsaraDB RDS for MySQL instance without interrupting the services of applications. This section describes how to migrate data from a local database with a private IP address to an ApsaraDB RDS for MySQL instance.

Background

DTS allows you to perform schema migration, full data migration, and incremental data migration on MySQL databases.

- **Schema migration**

  DTS migrates the schema definition of a local database to the destination instance. Currently, DTS supports schema migration for the following objects: tables, views, triggers, stored procedures, and storage functions.

- **Full data migration**

  DTS migrates all existing data of objects from a local database to the destination instance. If you also select incremental data migration, non-transaction tables without primary keys are locked during the full data migration process. Data cannot be written to these locked tables, and the locking duration depends on the data volume of the tables. The locks are released only after these tables are migrated. In this way, data consistency is guaranteed.

- **Incremental data migration**

  In incremental data migration, data changes made during the migration are updated to the destination instance. If DDL operations are performed during migration, the schema changes are not migrated to the destination instance.

Migration restrictions

Migrating data from a local database to an ApsaraDB RDS for MySQL instance is subject to the following restrictions:

- **DDL operations are not supported during migration.**
- **Event migration is not supported in schema migration.**
- **If you use the object name mapping feature when adding an object to be migrated, other objects associated with this object may fail to be migrated.**
- **When incremental data migration is selected, binlogging must be enabled and binlog_format must be set to ROW for the local MySQL instance. If the local MySQL version is 5.6, binlog_row_image must be set to FULL.**

Prerequisites

The ApsaraDB RDS for MySQL instance has been created, and a whitelist has been configured for it. For more information, see the "Set a whitelist" section of the *ApsaraDB for RDS User Guide* **.**

Prepare local data

Before the migration, create the migration accounts in the local database and the RDS for MySQL instance. You also need to create the database to be migrated in the RDS for MySQL instance, and grant the read and write permissions of the database to the migration account. *Table 13-4: Migration types and required permissions* **lists the permissions required by the migration accounts of the source and destination instances when different migration types are used.**

Table 13-4: Migration types and required permissions

Migration type	Schema migration	Full data migration	Incremental data migration
Local database	select	select	· select · replication slave · replication client
ApsaraDB RDS for MySQL instance	Read and write permissions	Read and write permissions	Read and write permissions

1. Run the following command to create a migration account in the local database:

   ```
   CREATE USER 'username'@'host' IDENTIFIED BY 'password';
   ```

   Parameters:

   - username: The migration account that you want to create.
   - host: The host from which you log on to the database by using the account. As a local user, you can use `localhost` to log on to the database. To log on from any other hosts, you can use the wildcard value `%`.
   - password: The logon password for the account.

   For example, if you want to create account `William` with password `Changme123` for logging on to the local database from any hosts, run the following command:

   ```
   CREATE USER 'William'@'%' IDENTIFIED BY 'Changme123';
   ```

2. **Grant permissions to the migration account in the local database.** *Table 13-4: Migration types and required permissions* **lists the permissions required for the migration account of the local database.**

```
GRANT privileges ON databasename.tablename TO 'username'@'host' WITH
 GRANT OPTION;
```

Parameters:

- **privileges: The operation permissions granted to the account, such as SELECT, INSERT, and UPDATE. If you want to grant all permissions to the account, set the value to `ALL`.**
- **databasename: The database name. If you want to grant all database permissions to the account, set the value to the wildcard value `*`.**
- **tablename: The table name. If you want to grant all table permissions to the account, set the value to the wildcard value `*`.**
- **username: The account to which you want to grant the permissions.**
- **host: The host from which the account is authorized to log on to the database. As a local user, you can use `localhost` to log on to the database. To log on from any other hosts, you can use the wildcard value `%`.**
- **WITH GRANT OPTION: Optional. This parameter enables the account to use the GRANT command.**

For example, if you want to grant all of the database and table permissions to account `William` and use the account to log on to the local database from any hosts, run the following command:

```
GRANT ALL ON *. * TO 'William'@'%';
```

Note:
If you want to perform incremental data migration, follow these steps to enable binlogging for the local database and configure this feature correctly.

3. **Run the following command to check whether binlogging has been enabled:**

   ```
   show global variables like "log_bin";
   ```

   If the query result is `log_bin=OFF`, binlogging has not been enabled for the local database. For synchronous migration of the incremental data generated in the migration process, modify the following parameters in configuration file *`my.cnf`*.

   ```
   log_bin=mysql_binbinlog_format=rowserver_id = integer greater than
   1binlog_row_image=full //When the local MySQL version is later than 5.
   6, this item must be set.
   ```

4. **After the parameters are set, run the following commands to restart the MySQL process:**

   ```
   $ Mysql_dir/bin/mysqladmin-u root-P Shutdown
   $ Mysql_dir/bin/maid &
   ```

   ***`mysql_dir`* is the installation directory of MySQL.**

Procedure

Perform migration after data preparation is completed.

1. *Log on to the DTS console*.
2. **In the left-side navigation pane, click Data Migration. On the page that appears, click Create Migration Task in the upper-right corner. In the Create DTS Instance dialog box that appears, create an instance as prompted.**

   Table 13-5: Parameters in the Create DTS Instance dialog box **describes the related parameters.**

   Table 13-5: Parameters in the Create DTS Instance dialog box

Parameter	Description
Function	The instance function. It is specified by the system. Current value: Data Migration.
Region	The region where the instance is located.
Created Instances	The number of instances to be created.

3. **After setting the parameters, click Create.**

4. **In the migration task list, find the instance that you created and click Configure Migration Task on the right. On the Create Migration Task page that appears, complete the configurations as prompted.**

   Table 13-6: Parameters for creating a migration task **describes the related parameters.**

   Table 13-6: Parameters for creating a migration task

Category	Parameter	Description
-	Task Name	DTS automatically generates a task name for each task by default. You can change the default name to an informative one for easy task identification.
Source database information	Instance Type	The type of the source instance. Select User-Created Database with Public IP Address.
-	Source Instance Region	The region where the source instance is located.
-	Database Type	The type of the database to be migrated. Select MySQL.
-	Hostname or IP Address	The connection address of the database to be migrated.
-	Port	The port number of the database to be migrated. The default port number for a MySQL database is 3306.
	Database Account	The account used to log on to the database to be migrated.
	Database Password	The password of the account used to log on to the database to be migrated.
Destination database information	Instance Type	The type of the destination instance. Select RDS Instance.
	Destination Instance Region	The region where the ApsaraDB RDS for MySQL instance is located. It is the same region as that of the source instance.
	RDS Instance ID	The ID of the ApsaraDB RDS for MySQL instance.

Category	Parameter	Description
	Database Account	The account used to log on to the ApsaraDB RDS for MySQL instance.
	Database Password	The password of the account used to log on to the ApsaraDB RDS for MySQL instance.

Note:
After configuring the source and destination databases, you can click Test Connection to test the connectivity.

5. After setting the parameters, click Set Whitelist and Next to go to the Migration Types and Tasks page.
6. Select migration types. Select the objects to be migrated in the Object to Be Migrated area, and click the right arrow to add the selected objects to the Selected area.

Note:
To modify the name of an object to be migrated in the destination database, move the pointer over the database to be modified in the Selected area. The Edit button is displayed.

7. Click Precheck.

Note:

- A precheck is required before you can start the migration task. A migration task can be started only after it passes the precheck.
- If the precheck fails, click the info icon corresponding to each failed check item to view the failure details, troubleshoot the faults, and re-run the precheck.
- After troubleshooting, select the task from the task list and restart the precheck.

8. After the precheck succeeds, you can start the migration task. After the task starts, you can check the migration status and progress on the Migration Tasks page.

Subsequent operations

The migration accounts have been granted the read and write permissions. For security considerations, we recommend that you delete the accounts from the local database and the ApsaraDB RDS for MySQL instance after the data migration.

13.3.5 Migrate data between RDS instances

This topic describes how to configure a task to use DTS for migrating data between two RDS instances.

DTS allows you to migrate data between two RDS instances. For storage engines that support incremental data migration, you can also migrate data to the destination RDS instance without stopping the services of the source RDS instance. DTS only supports migration between homogeneous databases. For example, you can migrate data between two ApsaraDB RDS for MySQL databases. However, migration between heterogeneous databases is not supported.

Permission requirements

When you use DTS to migrate data between RDS instances, the permissions required for the migration accounts vary with migration types. *Table 13-7: Migration types and required permissions* **lists the permissions required for the migration accounts of the source and destination instances when different migration types are used.**

Table 13-7: Migration types and required permissions

Migration type	Schema migration	Full data migration	Incremental data migration
Source RDS instance	Read/write permissions	Read/write permissions	Read/write permissions
Destination RDS instance	Read/write permissions	Read/write permissions	Read/write permissions

Procedure

This section describes how to use DTS for migrating data between two RDS instances. The source and destination RDS instances can be the same or different, indicating that you can use DTS to migrate data within an RDS instance or between two RDS instances.

Create an RDS instance database

If the destination database does not exist in the destination RDS instance, DTS automatically creates the database during data migration. However, in either of the following two cases, you must manually create the destination database in the RDS console before configuring a migration task:

- The database name does not meet the requirements of RDS: A database name must be 1 to 64 characters in length and can contain lowercase letters, digits, underscores (_), and hyphens (-). It must start with a letter and end with a letter or digit.
- The destination database has a different name from the source database.

In either case, you must create the destination database in the RDS console before configuring the migration task. For more information, see the "Create an RDS instance" section in the RDS User Guide.

Create a migration account

When configuring a migration task, you must provide the migration accounts of the source and destination RDS instances. For the permissions required for the migration accounts, see *Permission requirements*. If you have not created a migration account for your source or destination RDS instance, follow the procedure for creating an RDS instance account. You need to create migration accounts for the source and destination RDS instances, and grant the created accounts read/write permissions for the databases and tables to be migrated.

Configure a migration task

After all of the prerequisites are met, you can start to configure a migration task. To configure a migration task, follow these steps:

1. *Log on to the DTS console*.
2. In the left-side navigation pane, click Data Migration. On the page that appears, click Create Migration Task in the upper-right corner.
3. In the Create DTS Instances dialog box that appears, select a region, enter the number of data migration instances to be created, and click Create.
4. In the message that appears, click OK.
5. In the migration task list, find the migration task that you created and click Configure Migration Task.

6. **Configure the source and destination database instances, and click Set Whitelist and Next in the lower-right corner.**

   In this step, configure the migration task name, source RDS instance, and destination RDS instance. Parameters are described as follows:

   - **Task name**

     DTS automatically generates a task name for each task. Task names are not required to be unique. You can modify the task name as required. We recommend that you use an informative name for easy identification.
   - **Source instance information**
     - **Instance Type: Select RDS Instance.**
     - **RDS Instance ID: Select the ID of the source RDS instance. DTS supports RDS instances in classic networks and VPCs.**
     - **Database Name: Select the default database name that is used to connect to the source RDS instance.**
     - **Database Account: Enter the account that is used to access the source RDS instance.**
     - **Database Password: Enter the password of the account that is used to access the source RDS instance.**
   - **Destination instance information**
     - **Instance Type: Select RDS Instance.**
     - **Instance Region: Select the region where the destination instance resides.**
     - **RDS Instance ID: Select the ID of the destination RDS instance. DTS supports RDS instances in classic networks and VPCs.**
     - **Database Account: Enter the account that is used to access the destination RDS instance.**
     - **Database Password: Enter the password of the account that is used to access the destination RDS instance.**

7. **Select migration types and objects to be migrated.**

   In this step, you need to select migration types and objects to be migrated. Parameters are described as follows:

   - **Migration types**

     Available migration types are schema migration, full data migration, and incremental data migration.

     To perform a full data migration, select both Schema Migration and Full Data Migration for Migration Type.

     To perform a zero downtime migration, select Schema Migration, Full Data Migration, and Incremental Data Migration.

   - **Objects to be migrated**

     Select the objects to be migrated. Click the right arrow to add the selected objects to the Selected section on the right. An object to be migrated can be a database, table, or column.

     After objects are migrated to the destination database, the object names remain the same as those in the source database by default. If the object you migrate has different names in the source and destination instances, you must use the object name mapping feature provided by DTS. For more information about using this feature, see *Database, table, and column name mapping***.**

8. **Perform a precheck.**

   A precheck is required before you can start the migration task. A migration task can only be started after it passes the precheck.

   If the migration task fails the precheck, click the information icon correspond ing to each failed check item to view details, fix the issue, and run the precheck again.

   Click the info icon corresponding to each failed check item to view the cause and solution.

   After troubleshooting, select the task from the task list and perform the precheck again.

9. **Start the migration task.**

   After the migration task passes the precheck, you can start the migration task and check the migration status and progress in the task list.

   Incremental data migration is a dynamic synchronization process. We recommend that you verify the services in the destination database when data is consistent between the source and destination instances during incremental data migration. If the verification is successful, you can stop the migration task and switch services to the destination database.

   At this point, you have configured the task for migrating data between two RDS instances.

13.3.6 Migrate data from a local Oracle instance to an ApsaraDB RDS for MySQL instance

You can use DTS to migrate data from a local Oracle instance to an ApsaraDB RDS for MySQL instance. DTS supports schema migration, full data migration, and incremental data migration. You can use these three migration types in combination to migrate data from the Oracle instance to the destination instance without interrupting normal services of the source Oracle database. This section describes how to configure a task to use DTS to migrate data from an Oracle instance to an ApsaraDB RDS for MySQL instance without service interruptions.

Background

For data migration from an Oracle instance to an ApsaraDB RDS for MySQL instance, DTS supports schema migration, full data migration, and incremental data migration. The restrictions on each migration type are as follows:

- **Schema migration**

  DTS migrates the schema definitions of objects to the destination instance. Currently, DTS supports schema migration only for tables. For other objects such as views, synonyms, triggers, stored procedures, stored functions, packages, and user-defined data types, schema migration is not supported.

- **Full data migration**

  DTS migrates all existing data of objects from the source database to the destination ApsaraDB RDS for MySQL instance. If you perform only a full data migration, data changes to the source Oracle database during the migration

may not be migrated to the destination ApsaraDB RDS for MySQL instance. Therefore, if you only want to perform a full data migration without migrating the incremental data, we recommend that you stop writing data to the source Oracle instance during the migration to ensure data consistency.

- Incremental data migration

  During an incremental data migration, DTS polls and captures the redo logs generated by the source Oracle instance due to data changes. Then, DTS synchronizes the incremental data (or changed data) to the destination ApsaraDB RDS for MySQL instance in real time. Incremental data migration enables real-time data synchronization from the source Oracle instance to the destination ApsaraDB RDS for MySQL instance.

Permission requirements for migration

When you use DTS to migrate data from an Oracle instance to an ApsaraDB RDS for MySQL instance, the permissions required for the migration accounts vary with migration types. The following table lists the permissions required for the migration accounts of the source and destination instances when different migration types are used.

Migration type	Schema migration	Full data migration	Incremental data migration
Local Oracle instance	Schema owner	Schema owner	SYSDBA
Destination ApsaraDB RDS for MySQL instance	Read and write permissions on the database to be migrated	Read and write permissions on the database to be migrated	Read and write permissions on the database to be migrated

Prerequisites

- The version of the Oracle database to be migrated is 10g, 11g, or 12c.
- Supplemental logging has been enabled for the Oracle instance, and supplemental_log_data_pk and supplemental_log_data_ui have been enabled.
- Archive logging has been enabled for the Oracle instance, and archived logs can be accessed over a specific period.

Data type mappings

Oracle and MySQL have different data types. DTS needs to map the data types of the source and destination instances during schema migration. The following table lists the data type mappings between the source and destination instances.

Oracle data type	MySQL data type	Supported by DTS
varchar2(n [char/byte])	varchar(n)	Yes
nvarchar2[(n)]	national varchar[(n)]	Yes
char[(n [byte/char])]	char[(n)]	Yes
nchar[(n)]]	national char[(n)]	Yes
number[(p[,s])]	decimal[(p[,s])]	Yes
float(p)]	double	Yes
long	longtext	Yes
date	datetime	Yes
binary_float	decimal(65,8)	Yes
binary_double	double	Yes
timestamp[(fractional _seconds_precision)]	datetime[(fractional _seconds_precision)]	Yes
timestamp[(fractional _seconds_precision)]with local time zone	datetime[(fractional _seconds_precision)]	Yes
timestamp[(fractional _seconds_precision)]with local time zone	datetime[(fractional _seconds_precision)]	Yes
clob	longtext	Yes
nclob	longtext	Yes
blob	longblob	Yes
raw	varbinary(2000)	Yes
long raw	longblob	Yes
bfile	—	No
interval year(year_precision) to mongth	—	No

Oracle data type	MySQL data type	Supported by DTS
interval day(day_precision) to second[(fractional_seconds_precision)]	—	No

- For char(n), when the definition length n exceeds 255, DTS automatically converts the type to varchar(n).
- MySQL does not support the following data types in Oracle: bfile, interval year to month, and interval day to second. DTS does not convert these three data types during schema migration. You need to exclude columns of these three data types when you specify the objects to be migrated. If any of these three data types are included in the table to be migrated, the schema migration fails.
- The timestamp data type in MySQL does not contain time zone information, while the following data types in Oracle contain time zone information: timestamp with time zone and timestamp with local time zone. DTS converts these two types of data to the UTC time zone format before storing the data to the ApsaraDB RDS for MySQL instance during the migration.

SQL operations supported for synchronization

During incremental data migration, SQL operations that are supported for synchronization include:

- INSERT, DELETE, and UPDATE
- CREATE TABLE // Partitioned tables or tables with built-in functions are not supported.
- ALTER TABLE ADD COLUMN, ALTER TABLE DROP COLUMN, ALTER TABLE RENAME COLUMN, and ALTER TABLE ADD INDEX
- DROP TABLE
- RENAME TABLE, TRUNCATE TABLE, and CREATE INDEX

Create a migration account

When configuring a migration task, you must provide the migration accounts of the local Oracle instance and the destination ApsaraDB RDS for MySQL instance. For permissions required for the migration accounts, see the *Permission requirements for migration* section.

If you have not created a migration account for your source Oracle instance, follow the instructions in *Oracle GRANT Syntax* to create a migration account that meets the requirements.

For more information about how to create a migration account for the destination ApsaraDB RDS for MySQL instance and grant permissions to the account, see the " Create an account" section of the ApsaraDB for RDS User Guide.

Procedure

The following part describes how to configure a task to use DTS to migrate data from a local Oracle database to an ApsaraDB RDS for MySQL instance.

1. *Log on to the DTS console*.
2. **In the left-side navigation pane, click Data Migration. On the page that appears, click Create Migration Task in the upper-right corner. In the Create DTS Instance dialog box that appears, create an instance as prompted.**

   The following table describes the parameters in the Create DTS Instance dialog box.

   Table 13-8: Parameters in the Create DTS Instance dialog box

Parameter	Description
Function	The instance function. It is specified by the system. Current value: Data Migration.
Region	The region where the instance is located.
Created Instances	The number of instances to be created.

3. **In the migration task list, find the instance that you created and click Configure Migration Task on the right. On the Create Migration Task page that appears, complete the configurations as prompted.**

   Configurations:

   - **Task Name**

     DTS automatically generates a task name for each task. Task names do not have to be unique. You can modify the task name as required. We recommend that you use an informative name for easy task identification.
   - **Source instance information**
     - **Instance Type: Select User-Created Database with Public IP Address.**
     - **Database Type: Select Oracle.**
     - **Hostname or IP address: Enter the address for accessing the Oracle instance . This address must allows access from public networks.**
     - **Port: Specify the listening port of the Oracle instance.**
     - **Instance Type: Select Non-RAC Instance or RAC Instance based on the local Oracle data type.**
     - **SID: Specify the SID of the Oracle instance.**

       Note:
       This parameter is displayed only when you select Non-RAC Instance as the instance type.
     - **Service Name: Specify the server name of the instance.**

       Note:

This parameter is displayed only when you select RAC Instance as the instance type.

- **Database Account: Enter the account used to connect to the Oracle instance.**
- **Database password: Enter the password of the account used to connect to the Oracle instance.**

· **Destination instance information**

- **Instance Type: Select RDS Instance.**
- **RDS Instance ID: Select the ID of the destination ApsaraDB RDS for MySQL instance. DTS supports RDS instances in classic networks and VPCs.**
- **Database Account: Enter the account used to connect to the ApsaraDB RDS for MySQL instance.**
- **Database Password: Enter the password of the account used to connect to the ApsaraDB RDS for MySQL instance.**

After completing the configurations, click Set Whitelist and Next in the lower-right corner to configure the whitelist. In this step, DTS adds the IP address of the DTS server to the whitelist of the destination ApsaraDB RDS for MySQL instance. This prevents the connection issue where the DTS server cannot connect to the destination ApsaraDB RDS for MySQL instance for data migration.

4. **Select migration types and objects to be migrated.**

· **Migration types**

- **Schema migration**
- **Full data migration**
- **Incremental data migration**

Note:

- **To perform a zero downtime migration, select schema migration, full data migration, and incremental data migration.**
- **To perform only full data migration, select schema migration and full data migration.**

· **Objects to be migrated**

Select the objects to be migrated. An object to be migrated can be a database, a table, or a column. By default, after an object is migrated to the ApsaraDB RDS for MySQL instance, the object name remains the same as that in the local

Oracle instance. If the object you migrate has different names in the source and destination instances, use the object name mapping feature provided by DTS. For more information, see *Database, table, and column name mapping*.

After selecting migration types and objects to be migrated, click Precheck in the lower-right corner to proceed with a precheck.

5. **Perform a precheck.**

   A precheck is required before you can start the migration task. A migration task can be started only after it passes the precheck. For more information about the precheck items, see *Precheck items*.

   If the precheck fails, click the info icon corresponding to each failed check item to view the failure details, troubleshoot the faults, and re-run the precheck.

   After troubleshooting, select the task from the task list and restart the precheck.

6. **Start the migration task.**

   After the precheck succeeds, you can start the migration task and check the migration status and progress in the task list.

   When a migration task enters the incremental data migration stage, it does not automatically stop. Incremental data written to the Oracle instance is automatically synchronized to the destination ApsaraDB RDS for MySQL instance. Incremental data migration is a dynamic synchronization process. We recommend that you verify the services in the destination database when data is consistent between the source and destination instances during the incrementa l data migration. If the verification is successful, you can stop the migration task and switch the services to the destination database.

   Now, you have configured the task for migrating data from a local Oracle instance to an ApsaraDB RDS for MySQL instance.

Subsequent operations

The migration accounts have been granted the read and write permissions. For security considerations, we recommend that you delete the accounts from the local database and the ApsaraDB RDS for MySQL instance after the data migration.

13.3.7 Migrate data from an ApsaraDB RDS for MySQL instance to an on-premises Oracle database

You can use DTS to migrate data from an ApsaraDB RDS for MySQL instance to an on-premises Oracle database. DTS supports full data migration and incremental data migration to ensure that the source ApsaraDB RDS for MySQL instance can still provide services during data migration. This topic describes how to configure a task to use DTS for migrating data from an ApsaraDB RDS for MySQL instance to an on-premises Oracle database without disrupting your businesses.

Prerequisites

Schema migration is not supported when you migrate data from an ApsaraDB RDS for MySQL instance to an on-premises Oracle database. Before migration, you must create a destination Oracle database with the same schema as the source database in the ApsaraDB RDS for MySQL instance.

Permission requirements

When you use DTS to migrate data from an ApsaraDB RDS for MySQL instance to an on-premises Oracle database, the permissions required for the migration accounts vary with migration types. The following table lists the permissions required for the migration accounts of the source and destination instances when different migration types are used.

Migration type	Full data migration	Incremental data migration
Source ApsaraDB RDS for MySQL instance	Read/write permissions	Read/write permissions
Destination Oracle instance	Database read/write permissions	Database read/write permissions

Procedure

1. *Log on to the DTS console*.

2. **In the left-side navigation pane, click Data Migration. On the page that appears, click Create Migration Task in the upper-right corner. In the Create DTS Instances dialog box that appears, create an instance as prompted.**

   The following table describes the parameter settings.

   Table 13-9: Parameter descriptions

Parameter	Description
Feature	The feature specified by the system. In this case, the value is Data Migration.
Region	The region where the instance resides.
Instances to Create	The number of instances to be created.

3. **In the migration task list, find the instance that you created and click Configure Migration Task.**
4. **(Optional) Create a name for the task.**

   DTS automatically generates a name for every task. Task names are not required to be unique. You can modify the task name. We recommend that you use an informative name for easy identification.
5. **Enter information about the source and destination databases. The following table describes the parameter settings.**

Database type	Parameter	Description
Source database information	Instance Type	Instance Type: Select RDS Instance as the type of the source instance.
	Instance Region	Select the region where the source instance resides.
	RDS Instance ID	Select the ID of the source database.
	Database Account	Enter an account that has read/write permissions on the source database.
	Database Password	Enter the password of the source database account.
Destination database information	Instance Type	Instance Type: Select User-Created Database with Public IP Address as the type of the destination database.
	Instance Region	Select the region where the destination instance resides.

Database type	Parameter	Description
	Database Type	Select Oracle.
	Hostname or IP Address	Enter the hostname or IP address for accessing the Oracle instance.
	Port Number	The default value is 1521.
	Instance Type	Select Non-RAC Instance or RAC Instance based on the Oracle instance type.
	SID	Specify the system ID (SID) of the Oracle database.
	Database Account	Enter an account that has read/write permissions on the destination database.
	Database Password	Enter the password of the destination database account.

6. Click Test Connectivity and confirm that the test results for both the source and destination databases are Passed.
7. Click Set Whitelist and Next in the lower-right corner of the page.
8. Select a migration type. In the Available section, select the source database and click the right arrow to add the database to the Selected section.
   - To migrate data without stopping the running services, you must select Full Data Migration and Incremental Data Migration.
   - To perform full data migration, select Full Data Migration.
9. Click Precheck and wait until the precheck is complete.

   Note:
   For more information about the precheck items, see *Precheck items*. If the migration task fails the precheck, click the info icon corresponding to each failed check item to view details, fix the issue, and run the precheck again.
10. Click Next to start the migration task.

    After the task is started, you can check the migration status and progress in the task list.

    Note:

A migration task does not automatically stop after it reaches the incremental data migration stage. Incremental data written to the ApsaraDB RDS for MySQL instance is automatically synchronized to the destination Oracle instance. Incremental data migration is a dynamic synchronization process. We recommend that you verify the services running on the destination database when data is consistent between the source and destination instances during incremental data migration. If the verification is successful, you can stop the migration task and switch services to the destination database.

11. **At this point, you have configured the task for migrating data from an ApsaraDB RDS for MySQL instance to an on-premises Oracle database.**

13.3.8 Database, table, and column name mapping

This topic describes how to use the object name mapping feature when you configure a data migration task.

The data migration feature provided by DTS supports object name mapping. Objects to be migrated, such as databases, tables, and columns, can have different names in the source and destination instances.

Database name mapping

If a database you migrate has different names in the source and destination instances, you can map the database names by using the object name mapping feature of DTS.

You can configure the database name mapping feature in Step 2 "Select migration types and objects to be migrated" when you configure the migration task. Perform the following steps to configure the database name mapping feature:

1. **In the Selected area, move the pointer over the row of the object that requires database name mapping. The Edit button appears on the right.**

2. **Modify the database name.**

   If you want the database name to change to jiangliutest after the database is migrated to the destination instance, click Edit to open the Edit Database Name dialog box.

   In the Edit Database Name dialog box, modify the database name directly. The database is stored under the new name in the destination instance.

   Assume that the original database name is amptest.

   In the Edit Database Name dialog box, change amptest to jiangliutest, so that the database name changes to jiangliutest after the database is migrated to the destination instance.

   The database uses the new name in the destination instance.

Table name mapping

If a table you migrate has different names in the source and destination instances, you can map the table names by using the object name mapping feature of DTS.

If you want to use the table name mapping feature, do not select the entire database as the object to be migrated. Instead, select a specific table.

Besides tables, other schema objects such as views, stored procedures, stored functions, and synonyms are also available for object name mapping in the similar way.

You can configure the table name mapping feature in Step 2 "Select migration types and objects to be migrated" when you configure the migration task. Perform the following steps to configure the table name mapping feature:

1. **In the Selected area, move the pointer over the row of the object that requires table name mapping. The Edit button appears on the right.**

2. **Modify the table name.**

   If you want the table name to change from amptest to jiangliutest after the table is migrated to the destination instance, click Edit to open the Edit Table Name dialog box.

   In the Edit Table Name dialog box, modify the table name directly. The table is stored under the new name in the destination instance.

   Assume that the original table name is amptest.

   Change amptest to jiangliutest, so that the table name changes to jiangliutest after the table is migrated to the destination instance.

Column name mapping

If columns of a table that you migrate have different names in the source and destination instances, you can use the object name mapping feature provided by DTS.

You can configure the column name mapping feature in Step 2 "Select migration types and objects to be migrated" when you configure the migration task. If you want to modify the name of a column to be migrated, do not select the entire database as the object to be migrated. Instead, select the table that the column belongs to. Perform the following steps to configure the column name mapping feature:

1. **Assume that you want to change the name of a column in the sbtest1 table. In the Selected area, move the pointer over the row of the sbtest1 table. The Edit button appears on the right.**
2. **Click Edit to open the Edit Table Name dialog box.**

   Modify the column name. After the modification, the column is stored under the new name in the destination database.

   Now, you have configured the column name mapping feature.

13.3.9 Configure an SQL filter for filtering the data to be migrated

This section describes how to configure an SQL filter for filtering migration data when you create a migration task.

DTS allows you to configure an SQL filter to filter the table data to be migrated. The SQL filter applies only to the configured table. DTS filters the data in the table

of the source database based on this filter. Only data that meets this filter can be migrated to the destination database. This feature is applicable to multiple scenarios such as regular incremental data migration and table partitioning.

Functional restrictions

The SQL filter applies only to full data migration. If you select Incremental Data Migration as the migration type, the SQL filter does not apply.

Configure an SQL filter

You can configure an SQL filter in the Migration Types and Tasks step of migration task configuration.

If you want to configure an SQL filter for table migration, you must select a specific table instead of the entire database as the object to be migrated. The following part describes how to configure an SQL filter.

Configure an SQL filter

1. **In the Migration Types and Tasks step, move the pointer over the table for which you want to create an SQL filter in the Selected area. The Edit button appears.**
2. **Click Edit to configure a filter.**

Modify an SQL filter

Filters in DTS are the same as the standard SQL WHERE conditions for databases and support calculation and simple functions.

Enter an SQL filter in the text box as needed.

Now, you have configured an SQL filter.

13.3.10 Troubleshoot migration errors

DTS provides the feature of online troubleshooting in multiple stages to fix migration errors. These stages include:

- **Schema migration**

  DTS supports data migration between heterogeneous data sources. If you import data of unsupported types to the destination instance during a schema migration , the migration fails.

- **Full data migration**

  During full data migration, the migration task may fail because the destination RDS instance does not have sufficient space or required IP addresses have been deleted from the whitelist. In this case, you can modify the task configurations and then restart the task.

DTS provides the online troubleshooting feature that allows you to resume a failed task when an error occurs during migration. The following sections describe how to troubleshoot errors that occur during schema migration and full data migration.

Troubleshoot errors occurred during schema migration

If a schema migration task fails, the task status changes to Migration Failed and the Rectify button appears.

Click Rectify next to a failed object.

Click Rectify next to each failed object. A troubleshooting dialog box appears.

Modify the schema definition based on the cause of failure. Click Rectify after you complete the modification and re-import the modified definition to the destination instance.

If the error persists after you click Rectify, the cause of failure changes to Troubleshooting Failed and the cause of troubleshooting failure is displayed. You need to continue troubleshooting based on the cause of troubleshooting failure until the troubleshooting is successful.

The details page of the schema migration appears after troubleshooting is successful, and the status of the object changes to Finished.

The task resumes after issues with all objects are rectified. For example, the task resumes by proceeding to the full data migration stage.

Troubleshoot errors occurred during full data migration

DTS provides the troubleshooting and retry feature for the following causes of failures:

- **If you fail to connect to the source or destination database, retry the task after you ensure that the network connection is established.**
- **If a connection to the source or destination database times out, retry the task after you ensure that the network connection is established.**

- **If the destination RDS instance does not have sufficient space or the instance is locked, retry the task after you scale up the RDS instance or clean up the instance log space.**
- **If MyISAM of the source database is corrupted, retry the task after troublesho oting.**

For other circumstances, if full data migration fails, DTS only offers the Ignore option. You can ignore the failed object and continue the migration of other objects .

If a full data migration task fails, the status of the task changes to Migration Failed and the Rectify button appears.

When a migration task fails, click Rectify next to a failed object.

If you encounter the preceding failures and the migration tasks can be retried, troubleshoot the errors as prompted. Then, click the Retry button on the full data migration details page to continue the data transfer in the task.

For other causes of failures, DTS only supports the Ignore operation to ignore the full data migration of the object. After you click Ignore, data of this object is not migrated, but data of other objects is migrated to the destination instance.

13.4 Data synchronization

13.4.1 Create a real-time synchronization task

DTS provides a user-friendly real-time data synchronization feature. You can configure a subscription channel with only three steps. This section describes how

to use DTS to quickly create a synchronization task between two ApsaraDB RDS for MySQL instances for real-time synchronization of RDS incremental data.

Synchronization restrictions

- **Synchronization mode**

  Currently, DTS supports only the following modes for real-time synchronization between ApsaraDB RDS for MySQL instances:

  - **From A to B: unidirectional synchronization between two instances. The synchronized objects must be read-only in instance B. Otherwise, a synchronization channel exception may occur.**
  - **From A to B, C, and D: one-to-many distributed synchronization mode. This synchronization mode poses no restrictions on the number of destination RDS instances, but requires that the synchronized object be read-only in the destination instance to avoid synchronization channel exceptions.**
  - **From B, C, and D to A: many-to-one data convergence mode. In the many-to-one mode, the objects to be synchronized through each synchronization channel must be different to guarantee full synchronization.**

  DTS does not support the following modes:

  - **From A to B to C: cascading synchronization mode.**
  - **Between A and B: bidirectional synchronization mode between two instances.**

    If you select any other unsupported synchronization modes during the synchronization channel configuration, the complicated topologies check item in the precheck fails.

- **Incompatible triggers**

  When the object to be synchronized is an entire database and the database contains a trigger that updates the synchronization table, the synchronized data may be inconsistent.

  Suppose that the source instance contains table A and table B. Table A has a trigger that inserts a row of data into table B after the row of data is inserted into table A. The synchronization task synchronizes data from the source instance to the destination instance, where table A and table B are respectively represented as table A' and table B'. During synchronization, the destination instance continuously replicates a full amount of data from the source instance

, including table A and table B. If you insert a row (1) into table A, the trigger in table A inserts a row (2) into table B. The row you inserted (1) and the row inserted by the trigger (2) are synchronized respectively to table A' and table B ' in the destination instance. When the row you inserted (1) is updated to table A', the trigger in table A' inserts another row (2') in table B'. As a result, one row insertion (1) in the source table A triggers two row insertions (2 and 2') in the destination table B'. Therefore, the data between the source and destination instances is inconsistent.

To solve this problem, you must delete the trigger in the destination instance and synchronize table B from the source instance.

Prerequisites

Before configuring a synchronization task, make sure that the source and destination RDS instances exist. If the instances are unavailable, create them first.

Procedure

The following part describes the procedure for creating a synchronization channel between RDS instances:

1. *Log on to the DTS console*.
2. In the left-side navigation pane, click Data Synchronization.
3. On the Data Synchronization page, click Create Synchronization Task in the upper-right corner to configure a synchronization task.
4. In the dialog box that appears, set the parameters and click Create.

   Note:
   Currently, DTS only supports the following instance types: MySQL and MaxCompute.
5. In the message that appears, click OK.
6. On the page where synchronization tasks are listed, select the region to which the synchronization task you created belongs.
7. Find the synchronization task you created and click Configure Synchronization Channel.

8. In the Select Source and Destination Instances for Synchronization Channel step, configure the source and destination instances to be connected through the synchronization channel.

   Configure the following information:

   - Synchronization Task Name

     The synchronization task name does not have to be unique. We recommend that you use an informative name for easy task identification and management .

   - Source instance information
     - Instance Type: The type of the source instance. Select RDS Instance.
     - Instance Region: The region where the source instance is located.
     - Instance ID: The ID of the source instance. For a MaxCompute instance, set the value to a project name.
   - Destination instance information
     - Instance Type: The type of the destination instance. Select RDS Instance.
     - Instance Region: The region where the destination instance is located.
     - Instance ID: The ID of the destination instance. For a MaxCompute instance , set the value to a project name.

   > **Note:**
   > The source and destination RDS instances must be different. When you select a value for Instance ID, only the IDs of the ApsaraDB RDS for MySQL instances under the current logon account are displayed in the drop-down list.

   After completing the configuration, click Set Whitelist and Next.
9. Create a synchronization account and click Next.
10. Select the objects to be synchronized in the Source Database Objects area, and click the right arrow to add the selected objects to the Selected area. Click Next.

    DTS allows you to select the objects to be synchronized at the granularity down to table level. You can choose to synchronize certain databases or tables.

    If you select an entire database as the object to be synchronized, the schema change operations, such as CREATE TABLE and DROP VIEW operations,

performed on all the objects in the database are synchronized to the destination database.

If you select a table as the object to be synchronized, only the `DROP TABLE, ALTER TABLE, TRUNCATE TABLE, RENAME TABLE, CREATE INDEX, and DROP INDEX` operations performed on the table are synchronized to the destination database.

Note that the RENAME TABLE operation may result in data inconsistency between the source and destination instances. Suppose that only table A needs to be synchronized from the source instance. If you perform the `rename A to B` operation in the source instance during the synchronization, the subsequent operations performed on the renamed table B are not synchronized to the destination instance. To solve this problem, you can synchronize the entire source database where table A and table B are located.

11. Configure the initial synchronization.

    DTS first performs the initial synchronization when you start a synchronization channel. During initial synchronization, the schemas and data of the objects to be synchronized are replicated from the source instance to the destination instance. These schemas and data are then used as the baseline for subsequent incremental data synchronization.

    Two options are available for initial synchronization: Initial Schema Synchronization and Initial Full Data Synchronization. You must select both Initial Schema Synchronization and Initial Full Data Synchronization by default.

12. ClickPrecheck to start a precheck.

    After the precheck is successful, you can click Start to start the synchronization task.

    After the synchronization task is started, the synchronization task list is displayed. The newly started synchronization task is now in the Initial Synchronization state. The duration of the initial synchronization depends on the data size of the objects to be synchronized from the source instance. After initial synchronization is completed, the synchronization channel enters the Synchronizing state. Now, the synchronization channel between the source and destination instances is established.

13.4.2 Synchronize data between RDS instances in real time

This section describes how to configure a task to use DTS to synchronize data between two Real-time data synchronization between ApsaraDB RDS for MySQL instances under the same account ApsaraDB RDS for MySQL instances in real time.

Supported functions

Real-time data synchronization between ApsaraDB RDS for MySQL instances under the same account.

Synchronization restrictions

- **Synchronization mode**

  Currently, DTS supports only the following modes for real-time synchronization between ApsaraDB RDS for MySQL instances:

  - **From A to B: unidirectional synchronization between two instances. The synchronized objects must be read-only in instance B. Otherwise, a synchronization channel exception may occur.**
  - **From A to B, C, and D: one-to-many distributed synchronization mode. This synchronization mode poses no restrictions on the number of destination RDS instances, but requires that the synchronized object be read-only in the destination instance to avoid synchronization channel exceptions.**
  - **From B, C, and D to A: many-to-one data convergence mode. In the many-to-one mode, the objects to be synchronized through each synchronization channel must be different to guarantee full synchronization.**

  DTS does not support the following modes:

  - **From A to B to C: cascading synchronization mode.**
  - **Between A and B: bidirectional synchronization mode between two instances.**

    If you select any other unsupported synchronization modes during the synchronization channel configuration, the complicated topologies check item in the precheck fails.

- **Functional restrictions**
  - **Incompatible trigger**

    When the object to be synchronized is an entire database and the database contains a trigger that updates the table to be synchronized, the synchronized data may be inconsistent.

    Suppose that source instance A contains table A and table B. Table A has a trigger that inserts a row of data into table B after the row of data is inserted into table A. The synchronization task synchronizes data from the source instance to the destination instance, where table A and table B are respectively represented as table A' and table B'. During synchronization, the destination instance continuously replicates a full amount of data from the source instance, including table A and table B. If you insert a row (1) into table A, the trigger in table A inserts a row (2) into table B. The row you inserted (1) and the row inserted by the trigger (2) are synchronized respectively to table A' and table B' in the destination instance. When the row you inserted (1) is updated to table A', the trigger in table A' inserts another row (2') in table B'. As a result, one row insertion (1) in the source table A triggers two row insertions (2 and 2') in the destination table B'. Therefore, the data between the source and destination instances is inconsistent.

    To solve this problem, you must delete the trigger from the destination instance and synchronize table B from the source instance.
  - **Restrictions on the RENAME TABLE operation**

    The RENAME TABLE operation may cause data inconsistency between the source and destination instances. Suppose that only table A needs to be synchronized from the source instance. If you perform the `rename A to B` operation in the source instance during the synchronization, the subsequent operations performed on the renamed table B will not be synchronized to the destination instance. To solve this problem, you can synchronize the entire source database where table A and table B are located.

Prerequisites

Before configuring a synchronization task, make sure that the source and destination RDS instances exist. If the instances are unavailable, create them first.

Procedure

The following part describes the procedure for creating a synchronization channel between RDS instances:

1. *Log on to the DTS console*.
2. **In the left-side navigation pane, click Data Synchronization.**
3. **On the Data Synchronization page that appears, click Create Synchronization Task in the upper-right corner to configure a synchronization task.**
4. **In the dialog box that appears, set the parameters and click Create.**

   Note:
   Currently, DTS only supports the following instance types: MySQL and MaxCompute.
5. **In the message that appears, click OK.**
6. **On the page where synchronization tasks are listed, select the region to which the synchronization task you created belongs.**
7. **Find the synchronization task you created and click Configure Synchronization Channel.**

8. **In the Select Source and Destination Instances for Synchronization Channel step, configure the source and destination instances to be connected through the synchronization channel.**

   Configure the following information:

   - **Synchronization Task Name**

     The synchronization task name does not have to be unique. We recommend that you use an informative name for easy task identification and management .

   - **Source instance information**
     - **Instance Type: The type of the source instance. Select RDS Instance.**
     - **Instance Region: The region where the source instance is located.**
     - **Instance ID: The ID of the source instance. For a MaxCompute instance, set the value to a project name.**
   - **Destination instance information**
     - **Instance Type: The type of the destination instance. Select RDS Instance.**
     - **Instance Region: The region where the destination instance is located.**
     - **Instance ID: The ID of the destination instance. For a MaxCompute instance , set the value to a project name.**

   Note:
   The source and destination RDS instances must be different. When you select a value for Instance ID, only the IDs of the ApsaraDB RDS for MySQL instances under the current logon account are displayed in the drop-down list.

   After completing the configuration, click Set Whitelist and Next in the lower-right corner.

9. **Create a synchronization account and click Next.**

10. Select the objects to be synchronized in the Source Database Objects area, and click the right arrow to add the selected objects to the Selected area. Click Next.

    DTS allows you to select the objects to be synchronized at the granularity down to table level. You can choose to synchronize certain databases or tables.

    If you select an entire database, all schema change operations, such as CREATE TABLE and DROP VIEW, performed on all the objects in the database are synchronized to the destination database.

    If you select a table, only the `DROP TABLE, ALTER TABLE, TRUNCATE TABLE, RENAME TABLE`, `CREATE INDEX, and DROP INDEX` operations performed on this table are synchronized to the destination database.

    Note that the RENAME TABLE operation may result in data inconsistency between the source and destination instances. Suppose that only table A needs to be synchronized from the source instance. If you perform the `rename A to B` operation in the source instance during the synchronization, the subsequent operations performed on the renamed table B will not be synchronized to the destination instance. To solve this problem, you can synchronize the entire source database where table A and table B are located.

11. Configure the initial synchronization.

    DTS first performs the initial synchronization when you start a synchroniz ation channel. During initial synchronization, the existing schemas and data of the objects to be synchronized are replicated from the source instance to the destination instance. These schemas and data are then used as the baseline for subsequent incremental data synchronization.

    Two options are available for initial synchronization: Initial Schema Synchronization and Initial Full Data Synchronization. You must select both Initial Schema Synchronization and Initial Full Data Synchronization by default.

12. ClickPrecheck to start a precheck.

    After the precheck is successful, you can click Start to start the synchronization task.

    After the synchronization task is started, the synchronization task list is displayed. The newly started synchronization task is now in the Initial Synchronization state. The duration of initial synchronization depends on

the data size of the objects to be synchronized from the source instance. After initial synchronization is completed, the synchronization channel enters the Synchronizing state. Now, the synchronization channel between the source and destination instances is established.

13.4.3 Synchronize data from an RDS instance to a MaxCompute instance in real time

This topic describes how to use Data Transmission Service (DTS) to create a task for real-time data synchronization from an RDS instance to a MaxCompute instance. This facilitates real-time data analysis.

Feature

- **DTS supports real-time data synchronization from an ApsaraDB RDS for MySQL instance to a MaxCompute instance under the same Alibaba Cloud account.**
- **DTS supports RDS instances in the classic network and VPCs.**

Objects to be synchronized

DTS only supports the synchronization of tables. Other types of objects are not supported.

Synchronization process

The synchronization process consists of two stages:

1. **Initial full data synchronization.**

   In this stage, all data stored in the ApsaraDB RDS for MySQL instance is replicated to the MaxCompute instance. Data in each synchronized table is replicated and stored independently to a full data table in the MaxCompute instance. The default table name is <source table name>_base. For example , if the source table is named t1, then the destination full data table in the MaxCompute instance is named t1_dts_base. When configuring the synchroniz ation task, you can modify the name prefixes of full data tables.
2. **Incremental data synchronization.**

   In this stage, all the incremental data in the ApsaraDB RDS for MySQL instance is synchronized to the MaxCompute instance in real time. The incremental data is stored in incremental data tables, and each synchronized table corresponds to an incremental data table. The default name of an incremental data table stored

in MaxCompute is <source table name>_log. When configuring the synchronization task, you can modify the name prefixes of incremental log tables.

Incremental log tables store both the incremental data and specific metadata. *Table 13-10: Schema* defines the schema of an incremental log table.

Table 13-10: Schema

record_id	operation_flag	utc_timestamp	before_flag	after_flag	col1	…	colN
1	I	1476258462	N	Y	1	…	JustInsert
2	U	1476258463	Y	N	1	…	JustInsert
2	U	1476258463	N	Y	1	…	JustUpdate
3	D	1476258464	Y	N	1	…	JustUpdate

Where:

- record_id: the unique ID of an incremental log entry. This field auto-increments for each new log entry. If an UPDATE operation is performed on a data record, two log entries are generated for an INSERT operation and a DELETE operation, separately. The two log entries have the same record_id.
- operation_flag: the operation type of the incremental log entry.

  Valid values:

  - I: an INSERT operation.
  - D: a DELETE operation.
  - U: an UPDATE operation.
- dts_utc_timestamp: the operation timestamp of the incremental log. It is also the timestamp of the binary log file for the UPDATE record. The timestamp is in the UTC format.
- before_flag: indicates whether the column values that follow the incremental log entry are pre-update values. Valid values: Y and N. The value of before_flag is Y if the column values that follow the log entry are pre-update values. The value of before_flag is N if the column values are post-update values.
- after_flag: indicates whether the column values that follow the incremental log entry are post-update values. Valid values: Y and N. The value of after_flag

is N if the column values that follow the log entry are pre-update values. The value of after_flag is Y if the column values are post-update values.

For different operation types, before_flag and after_flag of an incremental log entry are defined as follows:

- INSERT operation

record_i	operation_flag	utc_timestamp	before_flag	after_flag	col1	…	colN
1	I	1476258462	N	Y	1	…	JustInsert

For an INSERT operation, the column values that follow an incremental log entry are the newly inserted record values, that is, post-update values. Therefore, the value of before_flag is N and the value of after_flag is Y.

- UPDATE operation

record_i	operation_flag	utc_timestamp	before_flag	after_flag	col1	…	colN
2	U	1476258463	Y	N	1	…	JustInsert
2	U	1476258463	N	Y	1	…	JustUpdate

When an UPDATE operation is performed, two incremental log entries are generated. The two incremental log entries have the same record_id, operation_flag, and dts_utc_timestamp.

The second log entry records the pre-update values, so the value of before_flag is Y and the value of after_flag is N.

The second log entry records the post-update values, so the value of before_flag is N and the value of after_flag is Y.

- DELETE operation

record_i	operation_flag	dts_utc_timestamp	before_flag	after_flag	col1	…	colN
3	D	1476258464	Y	N	1	…	JustUpdate

For a DELETE operation, the column values that follow an incremental log entry are the deleted record values, that is, pre-update values. Therefore, the value of before_flag is Y and the value of after_flag is N.

For each table synchronized from an RDS instance to a MaxCompute instance, a full data table and an incremental data table are generated in the MaxCompute instance. To retrieve the full data of a specific table at a specific time, you must merge the corresponding full data table and incremental data table in the MaxCompute instance. The merging procedure will be described later.

Configure a synchronization task

This section describes how to configure a task for synchronizing data from an RDS instance to a MaxCompute instance in real time.

1. *Log on to the DTS console*.
2. **In the left-side navigation pane, click Data Synchronization.**
3. **On the Data Synchronization page that appears, click Create Synchronization Task in the upper-right corner to configure a synchronization task.**
4. **In the dialog box that appears, set the parameters and click Create.**

   Note:
   DTS only supports the following instance types: MySQL, AnalyticDB, and MaxCompute.
5. **In the dialog box that appears, click OK.**
6. **On the Synchronization Tasks page, select the region where the synchronization instance resides.**
7. **Find the created synchronization task and click Configure Synchronization Channel in the Actions column.**

8. **In the Configure Source and Destination Instances in Synchronization step, configure the source and destination instances to be connected through the synchronization channel.**

   The parameters are described as follows:

   - **Synchronization Task Name**

     The synchronization task name is not required to be unique. We recommend that you use an informative name to help you identify and manage the synchronization channel.
   - **Source Instance Details**
     - **Instance Type: the type of the source RDS instance. Only ApsaraDB RDS for MySQL is supported. In this example, select RDS Instance.**
     - **Instance Region: the region where the source RDS instance resides.**
     - **Instance ID: the ID of the source RDS instance.**
   - **Destination Instance Details**
     - **Instance Type: the type of the destination instance. RDS for MySQL, MaxCompute (formerly ODPS), and DataHub are supported. When you configure a synchronization channel from an RDS instance to a MaxCompute instance, select MaxCompute as the destination instance type.**
     - **Instance Region: the region where the destination instance resides.**
     - **Instance ID: the ID of the destination instance.**

   After configuring the preceding parameters, click Set Whitelist and Next in the lower-right corner.

9. **Authorize the DTS synchronization account.**

   In this step, you need to grant the DTS synchronization account the write permission on the MaxCompute instance. This allows DTS to replicate data to the MaxCompute instance.

   Grant the DTS synchronization account the following permissions on a project in the MaxCompute instance:

   - **CreateTable**
   - **CreateInstance**
   - **CreateResource**
   - **CreateJob**
   - **List**

   To ensure the synchronization task stability, we recommend that you do not revoke the write permission during the synchronization process. Click Next to create a synchronization account.

   After the account is authorized, you can select the objects to be synchronized.

10. **Click Next to select the objects to be synchronized.**

    After the required permissions on the MaxCompute instance are granted to the DTS synchronization account, proceed with the initial synchronization configuration and select the tables to be synchronized.

    In this step, you need to configure the initial synchronization and select the tables to be synchronized. Where:

    a. **Initial synchronization**

       Two options are available for initial synchronization: Initial Schema Synchronization and Initial Full Data Synchronization.

       During initial schema synchronization, DTS creates a table that has the same schema as the table to be synchronized in the MaxCompute instance. During initial full data synchronization, DTS replicates all the existing data in the table to be synchronized to the MaxCompute instance. We recommend

that you select both Initial Schema Synchronization and Initial Full Data Synchronization when configuring the synchronization task.

b. Select the tables to be synchronized

You can select only tables as objects to be synchronized rather than the entire database. Each table to be synchronized corresponds to a full data table and an incremental data table in the MaxCompute instance. To modify the name of a table, click Edit next to the table in the Selected area to open the Edit Table Name dialog box.

11. Click Precheck to start a precheck.

After the precheck is successful, click Start Task to start the synchronization task.

After the synchronization task is started, it is displayed in the synchronization task list. The new synchronization task appears in the Performing Initial Sync state. The duration of the initial synchronization depends on the data volume of the objects to be synchronized in the source instance. After the initial synchronization is complete, the synchronization channel changes to the Synchronizing status. At this point, the synchronization channel between the source and destination instances is established.

When the synchronization channel is in the Synchronizing state, you can query the full data table and incremental data table in MaxCompute.

At this point, you have configured the task for synchronizing data from an RDS instance to a MaxCompute instance in real time.

Full data merging

This section describes how to retrieve the full data of a table at a specific time from the full data table and incremental data table in the MaxCompute instance. DTS supports full data merging by running SQL statements in MaxCompute.

You can run SQL statements in MaxCompute to merge the full data table and incremental data table to retrieve the full data at the time (t). The following code example shows the SQL statements to run in MaxCompute:

```
insert overwrite table result_storage_table
select col1,
       col2,
       colN
  from(
select row_number() over(partition by t.primary_key_column
```

```
 order by record_id desc, after_flag desc) as row_number, record_id,
operation_flag, after_flag, col1,col2,colN
  from(
select incr.record_id, incr.operation_flag, incr.after_flag, incr.col1
, incr.col2,incr.colN
  from table_log incr
 where utc_timestamp< timestmap
 union all
select 0 as record_id, 'I' as operation_flag, 'Y' as after_flag, base.
col1, base.col2,base.colN
  from table_base base) t) gt
where record_num=1
  and after_flag='Y'
```

The variables in the preceding code are described as follows:

- **result_storage_table: the name of the table that stores the result set of full data merging.**
- **col1, col2, colN: the column names of the synchronized table.**
- **primary_key_column: the name of the primary key column of the synchronized table.**
- **table_log: the name of the incremental data table.**
- **table_base: the name of the full data table.**
- **timestamp: the time when the full data will be merged.**

In the preceding example, the testdb_20161010_base table is the full data table that corresponds to the testdb table. The testdb_20161010_log table is the incremental data table that corresponds to the testdb table.

You can run the following SQL statements in MaxCompute to query the full data of the testdb table at the 1476263486 time:

```
insert overwrite table testdb_1476263486
select id,
       name
  from(
select row_number() over(partition by t.id
 order by record_id desc, after_flag desc) as row_number, record_id,
operation_flag, after_flag, id, name
  from(
select incr.record_id, incr.operation_flag, incr.after_flag, incr.id,
incr.name
  from testdb_20161010_log incr
 where utc_timestamp< 1476263486
 union all
select 0 as record_id, 'I' as operation_flag, 'Y' as after_flag, base.
id, base.name
  from testdb_20161010_base base) t) gt
 where gt.row_number= 1
```

```
    and gt.after_flag= 'Y' ;
```

You can also use DataWorks to add full data merging nodes before you perform further computing and analysis. After full data is merged, downstream computing and analysis nodes can be automatically triggered. You can set an interval for periodic offline data analysis.

At this point, you have configured the task for synchronizing data from an RDS instance to a MaxCompute instance in real time and merged the full data of tables.

13.4.4 Configure two-way data synchronization between RDS instances

13.4.4.1 Overview

DTS supports two-way real-time data synchronization between RDS instances on any two clouds. This section describes how to use DTS to create a two-way synchronization task between two ApsaraDB RDS for MySQL instances for active geo-redundancy, geo-disaster recovery, and other scenarios.

13.4.4.2 Supported synchronization statements

Two-way synchronization between ApsaraDB RDS for MySQL instances supports all DML updates (including INSERT, UPDATE, and DELETE) and the following DDL updates:

- **ALTER TABLE, ALTER VIEW, ALTER FUNCTION, and ALTER PROCEDURE**
- **CREATE DATABASE, CREATE SCHEMA, CREATE INDEX, CREATE TABLE, CREATE PROCEDURE, CREATE FUNCTION, CREATE TRIGGER, CREATE VIEW, and CREATE EVENT**
- **DROP FUNCTION, DROP EVENT, DROP INDEX, DROP PROCEDURE, DROP TABLE, DROP TRIGGER, and DROP VIEW**
- **RENAME TABLE and TRUNCATE TABLE**

 **Note:**

To ensure the stability of a two-way synchronization channel, you can synchronize DDL updates on the same table in only one direction.

For example, for two-way synchronization, you must enable DDL synchronization in either the A-to-B or B-to-A direction. If DDL synchronization is configured in one

direction, it is not supported in the reverse direction. You can only perform DML synchronization.

13.4.4.3 Detect and resolve conflicts

To ensure data consistency, for two-way synchronized instances, make sure that records with the same primary key, business primary key, or unique key are updated only on one of the instances. If you unexpectedly update a record with the same primary key, business primary key, or unique key on both instances that are two-way synchronized, a synchronization conflict occurs. To maximize the stability of two-way synchronized instances, DTS supports detecting and resolving data conflicts.

Considerations

During two-way synchronization, the system time of the source and destination instances may not be the same. Additionally, synchronization delays may occur . For these reasons, DTS cannot guarantee that its conflict detection mechanism can completely prevent data conflicts. You must refactor certain business logic to ensure that records of the same primary key, business primary key, or unique key are updated only on one of the instances that are two-way synchronized.

Supported conflict types

Currently, DTS supports detecting the following conflict types:

- **Uniqueness conflicts caused by INSERT operations**

  A uniqueness conflict occurs when the synchronization of an inserted row violates the unique constraint. For example, if two instances in two-way synchronization insert a record with the same primary key value at almost the same time, one of the inserted records fails to be synchronized because a record with the same primary key value already exists in the destination instance.
- **Inconsistent records caused by UPDATE operations**

  Update conflicts occur in the following scenarios:

  - **The records to be updated do not exist in the destination instance. If the records to be updated do not exist, DTS automatically changes the UPDATE**

operation to the INSERT operation and inserts these records to the destination instance. In this case, duplicate unique key values may occur.

- The primary keys or unique keys of the records to be updated conflict with each other.

- A DELETE operation is made on non-existent records

  A delete conflict occurs when the records to be deleted do not exist in the destination instance.

  In this case, DTS automatically ignores the DELETE operation regardless of the conflict resolution policy that you have configured.

Supported conflict resolution policies

For the preceding synchronization conflicts, DTS provides the following resolution policies. You can select a conflict resolution policy as required when configuring two-way synchronization.

- TaskFailed: The synchronization task reports an error and automatically exits the process in case of a conflict.

  When the synchronization encounters a conflict of the preceding types, the synchronization task reports an error and automatically exits the process. The task enters a failed state and you must manually resolve the conflict. This method is the default conflict resolution policy.

- Ignore: The records in the destination instance are used in case of a conflict.

  When the synchronization encounters a conflict of the preceding types, the synchronization task skips the current synchronization statement and continues the process. The records in the destination instance are used.

- Overwrite: The conflict records in the destination instance are overwritten in case of a conflict.

  When the synchronization encounters a conflict of the preceding types, the conflict records in the destination instance are overwritten.

13.4.4.4 Synchronization restrictions

This section describes the restrictions in cross-cloud data synchronization using DTS.

Restrictions in data sources

Currently, only ApsaraDB RDS for MySQL instances support two-way synchroniz ation. Other heterogeneous data sources do not support two-way synchronization.

The destination instance cannot be an RDS instance that runs in standard access mode and has only a public network address.

Restrictions in synchronization architecture

Currently, DTS only supports two-way synchronization between two ApsaraDB RDS for MySQL instances. Two-way synchronization between more than two instances is not supported.

Feature restrictions

- **Incompatible with triggers**

  When you synchronize an entire database and the database contains a trigger that updates the synchronization table, the synchronized data may be inconsiste nt.

  For example, the object to be synchronized is database A that contains table a and table b. Table a has a trigger that inserts a row to table b after the row is inserted to table a. In this case, if an INSERT operation is performed on table a in the source instance during synchronization, the data in table b is inconsistent between the source and destination instances.

  To resolve this problem, you must delete the trigger in the destination instance, so that the data in table b is only synchronized from the source instance.

- **Restrictions in the RENAME TABLE operation**

  The RENAME TABLE operation may result in inconsistent synchronization data. For example, if the object to be synchronized only includes table a and the rename a to b command is executed in the source instance during synchronization, subsequent operations to the renamed table b are not synchronized to the destination database. To solve this problem, you can synchronize the entire database where table a and table b are stored.

- **Restrictions in DDL synchronization direction**

  To ensure the stability of a two-way synchronization channel, you can synchronize DDL updates on the same table in only one direction. For example, in A-to-B and B-to-A synchronization, you can implement DDL synchronization in either the A-to-B or B-to-A direction. If DDL synchronization is configured in one direction, it is not supported in the reverse direction.

13.4.4.5 Configure two-way data synchronization between RDS instances across IDCs

This topic describes how to configure two-way data synchronization between RDS instances across IDCs.

Prerequisites

To configure a synchronization task, ensure that the source and destination ApsaraDB RDS for MySQL instances are available for two-way synchronization. You must first create the required instances if they do not exist.

Procedure

1. *Log on to the DTS console*.
2. **In the left-side navigation pane, click Data Synchronization.**
3. **On the Data Synchronization page, click Create Synchronization Task in the upper-right corner to start the task configuration.**

   Note:

   Source Instance Region: Select the region where the source RDS instance resides.

   Source Instance Type: Select the type of the source instance. In this example, select MySQL.

   Destination Instance Region: Select the region where the destination RDS instance resides.

   Destination Instance Type: Select the type of the destination instance. In this example, select MySQL.

   Sync Mode: Select the synchronization mode. In this example, select Two-Way Synchronization.

Instances to Create: Set the number of instances that you want to create.

4. **After you configure the preceding information, click Create.**

   After you create a synchronization instance, go back to the Synchronization Tasks page. The new synchronization instance is in the Not Configured state and contains two synchronization tasks. You can configure two-way synchronization for the tasks.

5. **Find the required synchronization task and click Set Sync Channel.**
6. **Configure the required information for connecting to the synchronization channel.**

   Parameters are described as follows:

   - **Synchronization task name**

     The synchronization task name is not required to be unique. We recommend that you set an informative name to help you easily identify and manage the synchronization channel.

   - **RDS instance ID of the synchronization task**

     You must specify the ID of the Apsara Stack tenant account to which the destination RDS instance belongs. You can then select an RDS instance ID from the Instance ID drop-down list.

     After you complete the preceding configurations, click Set Whitelist and Next to configure the RDS instance whitelists.

7. **Configure the RDS instance whitelists.**

   In this step, add the IP addresses of the DTS servers to the whitelists of the source and destination RDS instances. This helps you avoid failure in creating a synchronization task when the DTS server cannot connect to the RDS instances because of the whitelist mechanism.

   We recommend that you do not remove the server IP addresses from the whitelists of the RDS instances. This ensures the stability of the synchronization task.

   Click Next to create a synchronization account.

8. Create a synchronization account for connecting to the destination database.

   In this example, create a synchronization account named dtssyncwriter in the destination RDS instance. During the synchronization, the account cannot be deleted. Otherwise, an interruption occurs.

9. Configure the objects to be synchronized and the synchronization policies.

   After you create a synchronization account for connecting to the destination RDS instance, you can start configuring the objects and synchronization policies.

   - Exclude DDL Statements

     This field determines whether to synchronize DDL statements in a specific direction. To include DDL statements, select No. To exclude DDL statements, select Yes. After you select No, the same table does not support synchronizing DDL statements in the other direction.

   - DML Statements for Synchronization

     This field determines the DML statements to be synchronized. By default, the INSERT, UPDATE, and DELETE statements are selected. You can select the DML statement types based on your business requirements.

   - Conflict Resolution Policy

     This field determines the resolution policy in case of a synchronization conflict. By default, TaskFailed is selected. You can select a conflict resolution policy based on your business requirements.

     For example, if Node A is the primary business center and Node B is a secondary business center, you must give the priority to Node A. Specifically, you need to set the conflict resolution policy in the A-to-B direction to Overwrite and that in the B-to-A direction to Ignore.

   - Objects to Be Synchronized

     The objects to be synchronized in real time include databases and tables.

     If you select an entire database, all schema update operations (such as CREATE TABLE and DROP VIEW) performed on all the objects in the database are synchronized to the destination database.

     If you select a table, only the DROP TABLE, ALTER TABLE, TRUNCATE TABLE, RENAME TABLE, CREATE INDEX, and DROP INDEX operations to this table are synchronized to the destination database.

10. Configure initial synchronization.

    Initial synchronization is the first step to start the synchronization channel. It synchronizes the schema and data of the objects to be synchronized in the source instance to the destination instance. The schema and data are used as the baseline data for subsequent incremental data synchronization.

    Initial synchronization includes Initial Schema Synchronization and Initial Full Data Synchronization. You must select both Initial Schema Synchronization and Initial Full Data Synchronization by default.

    If some tables to be synchronized in one direction are also included in the objects to be migrated in the other direction, these tables do not go through the initial synchronization process.

11. Precheck.

    After you complete the preceding configurations, perform the precheck before starting the synchronization task.

    After the precheck is passed, click Start to start the synchronization task.

    After the synchronization task is started, the synchronization task list is displayed. The newly started synchronization task changes to the Performing Initial Sync state. The duration of the initial synchronization depends on the data volume of the objects to be synchronized in the source instance. After initial synchronization, the synchronization channel changes to the Synchronizing state and the synchronization channel between the source and destination instances is established.

    After the synchronization task is configured in this direction, the source and destination RDS instances of the synchronization task in the other direction are fixed and cannot be changed.

12. After completing the synchronization task configurations in one direction, you can configure the synchronization task in the other direction. For more information about the steps, see step 6 to step 12 in the preceding section.

13.4.5 Troubleshoot precheck failures

Before a real-time synchronization channel is started, a precheck is performed. This topic describes the precheck items and how to troubleshoot precheck failures.

Source database connectivity

- **Description**

  This item checks the connectivity between the DTS server and the source RDS instance. DTS creates a connection to the source RDS instance by using the JDBC protocol. If the connection fails, the precheck fails.
- **Cause of failure**
  - **DTS does not support real-time synchronization between RDS instances in the region where the source instance resides.**
  - **The source instance account or password is incorrect.**
- **Solution**

  Submit a ticket and contact Alibaba Cloud technical support.

Destination database connectivity

- **Description**

  This item checks the connectivity between the DTS server and the destination RDS instance. DTS creates a connection to the destination RDS instance by using the JDBC protocol. If the connection fails, the precheck fails.
- **Cause of failure**
  - **DTS does not support real-time synchronization of RDS instances in the region where the destination instance resides.**
  - **The destination instance account or password is incorrect.**
- **Solution**

  Submit a ticket and contact Alibaba Cloud technical support.

Source database version

- **Description**

  This item checks whether:

  1. **The version of the source RDS instance is supported by the real-time synchronization feature.**
  2. **The version of the destination RDS instance is the same as the version of the source RDS instance.**

- **Cause of failure**

  - **The version of the source RDS instance is earlier than the versions supported by DTS. The source instance version must be MySQL 5.1, 5.5, or 5.6 for a real-time synchronization task.**
  - **The version of the destination RDS instance is earlier than the version of the source RDS instance.**

- **Solution**

  - **If the version of the source RDS instance is earlier than the versions supported by DTS, upgrade the source RDS instance to MySQL 5.6 in the RDS console. Then, re-create the synchronization channel.**
  - **If the version of the destination RDS instance is earlier than the version of the source RDS instance, upgrade the destination RDS instance to MySQL 5.6 in the RDS console. Then, re-create the synchronization channel.**

Database existence

This item checks whether the database to be synchronized already exists in the destination instance. If the database to be synchronized does not exist in the destination instance, DTS automatically creates a database. However, DTS fails to create the database and reports a failure under the following circumstances:

- **The database name contains characters other than lowercase letters, digits, underscores (_), and hyphens (-).**
- **The character set of the database is not UTF-8, GBK, Latin1, or UTF-8MB4.**
- **The migration account of the destination database does not have the read/write permission on the source database.**

  If the source database is an RDS instance, the precheck does not fail.

Source database permissions

This item checks whether the synchronization account of the source database has the required permissions. If the synchronization account does not have the required permissions, the precheck fails. If the source database is an RDS instance, the precheck does not fail.

Destination database permissions

- **Description**

  This item checks whether the synchronization account of the destination database has the required permissions. If the synchronization account does not have the required permissions, the precheck fails.
- **Cause of failure**
  - **DTS fails to create an account in the destination RDS instance.**
  - **DTS fails to grant the read/write permission to the synchronization account of the destination RDS instance.**
- **Solution**

  Submit a ticket and contact Alibaba Cloud technical support.

Object name conflict

- **Description**

  This check item applies only when you have configured initial synchroniz ation for a synchronization channel. The item checks whether an object to be synchronized has the same name as an object in the destination RDS instance.
- **Cause of failure**

  If an object in the destination RDS instance has the same name as the object to be synchronized, the precheck fails.
- **Solution**
  - **Remove the object that has the same name as the object to be synchronized from the destination database.**
  - **Then, re-create a synchronization channel. Two initial synchronization options are available: initial schema synchronization and initial full data synchronization.**

Source database server_id

This item checks whether server_id of the source database is set to an integer greater than or equal to 2. If the source database is an RDS instance, the precheck does not fail.

Whether binlogging is enabled for the source database

This item checks whether binlogging is enabled for the source database. If binlogging is not enabled for the source database, the precheck fails. If the source database is an RDS instance, the precheck does not fail.

Whether the binlog format is ROW in the source database

This item checks whether the binlog format is ROW in the source database. If the binlog format is not ROW in the source database, the precheck fails. If the source database is an RDS instance, the precheck does not fail.

Referential integrity constraint

- **Description**

  This item checks whether all the parent-child tables that have foreign key dependencies with the objects to be synchronized have been synchronized. This protects the integrity of foreign key constraints.
- **Cause of failure**

  Some of the objects to be synchronized are child tables with foreign key dependencies, but their parent tables have not been synchronized. This impairs the integrity of foreign key constraints.
- **Solution**

  The following solutions are available:

  - **Re-create the synchronization task and do not synchronize the child tables that failed the referential integrity constraint check.**
  - **Re-create the synchronization task and add the following tables to the list of objects to be synchronized: the parent tables for the child tables that failed the referential integrity constraint check.**
  - **Modify the source database and delete the foreign key dependencies of the child tables that failed the referential integrity constraint check. Then, re-create the synchronization task.**

Storage engine

- **Description**

  This item checks whether the objects to be synchronized use storage engines that are not supported by the real-time synchronization feature, such as Federated, MRG_MYISAM, and TokuDB.

- **Cause of failure**

  If a table to be synchronized uses the storage engine Federated, MRG_MyISAM, or TokuDB, the precheck fails.

- **Solution**

  Change the unsupported storage engines to InnoDB and re-create the synchronization task.

Character set

- **Description**

  This item checks whether the objects to be synchronized use character sets that are not supported by the real-time synchronization feature, such as the UCS-2 character set.

- **Cause of failure**

  If the character sets used by the objects to be synchronized are not supported by the real-time synchronization feature, the precheck fails.

- **Solution**

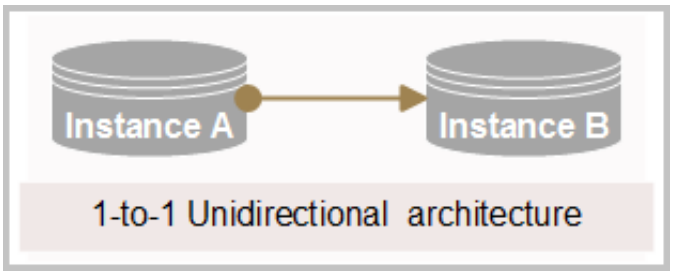
  Change the unsupported character sets to UTF-8, GBK, or Latin1. Then, re-create the synchronization task.

Complicated topologies

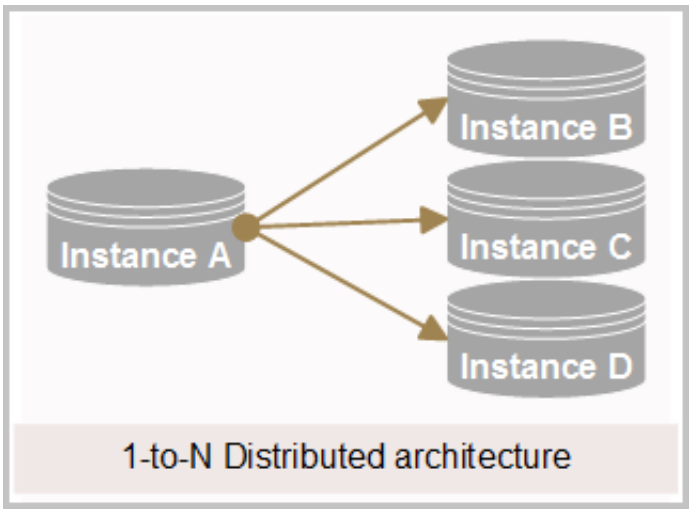
- **Description**

  This item checks whether the source and destination RDS instances of the synchronization task use unsupported synchronization modes. The real-time synchronization feature only supports two synchronization modes:

  - **One-to-one**

  - **One-to-many**

  The real-time synchronization feature does not support the many-to-one, cascading, and two-way synchronization modes.

- **Cause of failure**

  - **The destination RDS instance in the current synchronization task is the source RDS instance in another synchronization task.**
  - **Another synchronization task already runs on the destination RDS instance.**
  - **A migration task is in progress between the source and destination RDS instances, and the objects to be migrated in the migration task overlap the objects in the synchronization task to be created.**

- **Solution**
  - **If a synchronization task already runs between the source and destination RDS instances and this task is the same as the current task that you want to create for synchronizing new objects, do as follows: Modify the existing synchronization channel and do not create a new task. Then, add the expected objects to the list of objects to be synchronized.**
  - **If the synchronization channel conflicts with an existing migration task, wait until the migration task is complete before you can re-create the synchronization task.**
  - **If the new and original synchronization tasks constitute a cascading, two-way, or many-to-one synchronization mode, the mode is not supported for the moment.**

MySQL old password format

This item checks whether the password used by the source instance is an old password. If the source database is an RDS instance, the precheck does not fail.

13.4.6 Check the synchronization performance

DTS provides trend charts on synchronization latency, TPS, and traffic, allowing you to check the running performance of synchronization tasks in real time.

1. *Log on to the DTS console*.
2. **In the left-side navigation pane, click Data Synchronization.**
3. **In the synchronization task list, click the ID of the synchronization task you want to check.**

   The task details page appears.
4. **On the task details page, click Synchronization Performance in the left-side navigation pane.**

5. **View the trend charts on synchronization performance.**

   DTS provides trend charts on synchronization latency, TPS, and traffic.

   - **Synchronization traffic: The data traffic that the data writing module pulls from the data pulling module per second in DTS. The unit is MB/s.**
   - **Synchronization TPS: The number of queries that DTS synchronizes to the destination RDS instance per second.**
   - **Synchronization latency: The difference between the timestamp of the latest synchronized data in the destination RDS instance and the current timestamp in the source RDS instance. The unit is seconds.**

13.4.7 Add objects to be synchronized

DTS allows you to dynamically modify the objects to be synchronized during the synchronization process. This section describes how to add objects to be synchronized during a synchronization process.

Restrictions on object modifications

You can modify the objects to be synchronized only when the synchronization task is in the Synchronizing or Synchronization Failed state.

Synchronization start time

After an object to be synchronized is added, the synchronization start time depends on whether initial synchronization has been configured for the synchronization task.

- **If initial synchronization has been configured for the synchronization task, DTS performs initial synchronization on the new object before starting incremental synchronization.**
- **If initial synchronization has not been configured for the synchronization task, DTS starts synchronization immediately after incremental data is generated for the objects to be synchronized on the source instance.**

Procedure

1. *Log on to the DTS console*.
2. **In the left-side navigation pane, click Data Synchronization.**
3. **Find the synchronization task to be modified, and click View More > Modify Object to Be Synchronized to modify the objects to be synchronized.**

4. On the Select Object to Be Synchronized page, add objects to be synchronized as needed.
5. Click Next to start a precheck.

   After the precheck is successful, click Start.

   After an object to be synchronized is added, if the synchronization task requires initial synchronization, the task state changes from Synchronizing to Synchronizing (Initial synchronization is being performed on the new object...). At this time, the backend restarts the synchronization channel and the synchronization latency changes to -1 second. After the synchronization channel is restarted, the synchronization latency and speed return to normal.

   In the synchronization task list, you can click View Details to view the initial synchronization progress of the new objects. After the initial synchronization on the new objects is completed, the synchronization task returns to the Synchronizing state.

13.4.8 Remove objects to be synchronized

DTS allows you to dynamically modify the objects to be synchronized during the synchronization process. This section describes how to remove objects to be synchronized during a synchronization task.

Restrictions on object modifications

You can modify the objects to be synchronized only when the synchronization task is in the Synchronizing or Synchronization Failed state.

Procedure

1. *Log on to the DTS console*.
2. In the left-side navigation pane, click Data Synchronization.
3. Find the synchronization task to be modified, and click View More > Modify Object to Be Synchronized to modify the objects to be synchronized.
4. Click Configure Synchronization Instance in the Actions column corresponding to a synchronization task. On the Select Object to Be Synchronized page, remove objects to be synchronized as required.

   Now, you have deleted certain objects to be synchronized.

13.5 Change tracking

13.5.1 Overview

Change tracking is a feature provided by DTS that allows you to track data changes in ApsaraDB RDS for MySQL instances in real time. With this feature, you can complete lightweight cache update, asynchronous service decoupling, and real-time synchronization of data by using the Extract, Transform, Load (ETL) logic.

Objects for change tracking

Objects for change tracking include databases and tables. You can specify one or more tables for which you want to track data changes.

In change tracking, incremental data includes data manipulation language (DML) operations and data definition language (DDL) operations. When you configure change tracking, you must select operation types.

Change tracking channels

A change tracking channel is the basic unit of incremental data tracking and consumption. To track data changes in an RDS instance, you must create a change tracking channel in the DTS console for the RDS instance. The change tracking channel pulls incremental data from the RDS instance in real time and locally stores the data. You can use the DTS SDK to consume incremental data in the channel. You can also create, manage, or delete change tracking channels in the DTS console.

13.5.2 Create an RDS change tracking channel

Change tracking is a feature provided by DTS that allows you to track data changes in RDS instances in real time. With this feature, you can complete lightweight cache update, asynchronous service decoupling, and real-time synchronization of data by using the Extract, Transform, Load (ETL) logic.

Prerequisites

Limits on change tracking are listed as follows:

- **The change tracking feature only applies to ApsaraDB RDS for MySQL instances.**
- **The binlog_row_image value of MySQL 5.6 binlog must be full.**
- **Only the InnoDB and MyISAM storage engines are supported.**

- Only the following MySQL character sets are supported: latin1, gbk, utf8, utf8mb4, and binary.

Context

To use the change tracking feature to consume the incremental data of an RDS instance in real time, follow these two steps:

1. Create a change tracking channel for the RDS instance in the DTS console.
2. Use the SDK provided by DTS to access the change tracking channel and consume the incremental data in real time.

This topic describes how to create a change tracking channel in the DTS console. You can create a change tracking channel with only three steps. For more information about how to manage a change tracking channel and use the SDK, see the DTS Product Manual.

The following section describes the procedure for creating a change tracking channel.

Procedure

1. *Log on to the DTS console*.
2. In the left-side navigation pane, click Change Tracking.
3. On the Change Tracking page, click Create Change Tracking Task.
4. In the Create DTS instances dialog box that appears, select a region, enter the number of change tracking channels to be created, and click Create.
5. In the message that appears, click OK.
6. In the change tracking channel list, find the change tracking channel that you created, and click Set Channel in the Actions column.
7. Configure the RDS instance for change tracking.

   In this step, you need to configure the name of the change tracking channel and the ID of the RDS instance. Parameters are described as follows:

   - Task Name: the alias of the change tracking channel. It is not required to be unique. By default, DTS automatically generates a name for each change

tracking channel. You can set the name to an informative one for easy identification of the channel.

- RDS Instance ID: the ID of the RDS instance that contains the incremental data you want to consume.

  After completing the configuration, click Set Whitelist and Next in the lower-right corner.

8. Select the data type for the change tracking channel. Then, select the required objects in the left-side section, and click the right arrow to add the selected objects to the right-side Selected section.

   In this step, select the data types and objects required for change tracking. Parameters are described as follows:

   - Required Data Types

     DTS provides two types of data changes that can be tracked: Data Updates and Schema Updates. Data updates refer to any data changes made by DML operations, such as INSERT, DELETE, and UPDATE operations. Schema updates refer to the schema changes made by DDL operations, such as CREATE TABLE, DROP TABLE, and ALTER TABLE operations.

     If you select Schema Updates, DTS pulls all schema updates in the RDS instance. If you only want the schema updates made by certain DDL operations, you can set filters when you use the DTS SDK to consume data.

   - Objects for change tracking

     DTS allows you to select databases and tables as the objects for change tracking. You can track data changes to specific databases or tables.

9. Click Save and Precheck in the lower-right corner.

   After the precheck is passed, DTS starts the change tracking channel.

   DTS requires about one minute to perform initial change tracking.

   At this point, you have configured the change tracking channel.

13.5.3 Change consumption checkpoints

This section describes how to change consumption checkpoints in the DTS console.

Context

DTS allows you to change consumption checkpoints at any time during the consumption process. After a consumption checkpoint is changed, only data generated after the new consumption checkpoint can be pulled by the downstream SDKs as incremental data. The new consumption checkpoint must be within the data range of the subscription channel. Currently, you can change consumption checkpoints only in the DTS console, and cannot specify consumption checkpoints in the SDK.

Procedure

1. Stop the SDK consumption process.

   Before you change a consumption checkpoint, make sure that all the downstream SDKs connected to the subscription channel have been stopped. You can view the consumer sources (IP addresses) of the subscription channel in the DTS console to check whether all the downstream SDKs have been stopped.

   If the consumer sources are empty, all the downstream SDKs of the subscription channel have been stopped.

2. Change a consumption checkpoint.

   Currently, you can change a consumption checkpoint only in the DTS console.

   If you want to change a consumption checkpoint of the subscription channel, move the pointer over the checkpoint. A pen-like edit icon appears. Click the icon. The dialog box for changing the consumption checkpoint appears.

   > **Note:**
   > The consumption checkpoints configured here must be within the range of the current data tunnel.

3. Restart the SDK consumption process.

   After the consumption checkpoint is changed, restart the local SDK consumption process. Then the SDK subscribes to the incremental data from the new consumption checkpoint.

13.5.4 Modify objects for change tracking

This section describes how to modify objects for change tracking in the DTS console.

Context

DTS allows you to add or remove objects for change tracking in the consumption process. After you add an object, the change tracking channel pulls the incremental data of the new object from the time when the modification takes effect. After you remove an object, the SDK no longer subscribes to the data of the removed object from the time when the modification takes effect.

Procedure

1. Go to the Change Tracking Tasks page to modify the objects for change tracking.

   You can only modify the objects for change tracking in the DTS console.

   Find the change tracking channel for which you want to modify the required objects. Click View More in the Actions column, and select Modify Required Objects.

2. Modify objects for change tracking.

   After you click Modify Required Objects, the Select Required Objects page appears.

   On this page, you can add and remove objects for change tracking, or change the data type. After the objects for change tracking are modified, DTS starts a precheck.

   After the precheck is passed, click Start. Initial change tracking is performed on the change tracking channel.

   After initial change tracking is complete, the change tracking channel switches to the Active state and starts to work as expected. You can now use the SDK to track data changes.

13.5.5 Methods provided by SDK

SDK defines multiple classes. This topic describes the methods provided by these classes.

The DTS SDK is required for tracking and consuming incremental data.

Before using the SDK for data consumption, you must log on to the DTS console and create a change tracking channel for the RDS instance to which you want to subscribe.

After the change tracking channel is created, you can use the SDK to track data changes in real time.

- **DTS provides only the Java version of the SDK.**
- **The data in one change tracking channel can be consumed by only one SDK client . If multiple SDK clients are connected to the same change tracking channel, only one SDK process can pull the incremental data from the channel. If multiple downstream SDK clients need to subscribe to the incremental data in the same RDS instance, you must create a change tracking channel for each downstream SDK client.**

Methods of the RegionContex class

- **setAccessKey(AccessKey)**

  Specifies the AccessKey ID. Set the AccessKey parameter to the AccessKey ID of the account that subscribes to the change tracking channel.
- **setSecret(AccessKeySecret)**

  Specifies the AccessKey Secret. Set the AccessKeySecret parameter to the AccessKey Secret of the account that subscribes to the change tracking channel. You can go to the AccessKey page to create and obtain an AccessKey Secret.
- **setUsePublicIp(usePublicIp)**

  Specifies whether the server where the SDK is running subscribes to data changes over the public network. If the public network is used, set the usePublicIp parameter to True. If the public network is not used, set the usePublicIp parameter to False.

  Data changes can be tracked over internal networks. Before establishing a change tracking channel, the SDK communicates with the DTS control system over the Internet to obtain the physical connection address of the change tracking channel. If data changes must be tracked over internal networks, you need to attach a public IP address to the server where the SDK is deployed.

Methods of the ClusterClient class

- **void addConcurrentListener(ClusterListener arg0)**

  Adds downstream listeners. A listener can subscribe to incremental data in the change tracking channel only after the listener is added to a ClusterClient object.

  The ClusterListener arg0 parameter specifies an object of the ClusterListener class.

- **void askForGUID(String arg0)**

  Requests the incremental data from a specified change tracking channel. The String arg0 parameter specifies the ID of the change tracking channel. You need to obtain the ID in the DTS console.

- **List<ClusterListener> getConcurrentListeners()**

  Obtains the list of listeners in a ClusterClient object. The return type is List< ClusterListener>.

- **void start()**

  Starts the SDK client to subscribe to incremental data.

- **void stop()**

  Stops the SDK client to stop subscribing to incremental data. Data pulling and notification callback are performed in the same thread in the SDK. If the consumption code of the notify() method contains a function that prevents signal interruptions, the stop function may fail to terminate the client.

Methods of the ClusterListener class

void notify(List<ClusterMessage> arg0)

Defines the consumption of incremental data. After receiving data, the SDK client uses the notify() method to instruct a ClusterListener object to consume data. For example, the consumption mode in the SDK demo indicates that tracked data changes are displayed on the screen.

The input parameter type of this method is List<ClusterMessage>, in which ClusterMessage is the schema of tracked data changes. For more information, see *Methods of the ClusterMessage class*.

Methods of the ClusterMessage class

Each ClusterMessage object stores the data record of an RDS transaction. Each record is stored by using a Record object. This section introduces methods of the ClusterMessage class.

- **Record getRecord()**

  Obtains a change record from the ClusterMessage object. The change record indicates each log entry in the RDS binlog file, such as BEGIN, COMMIT, UPDATE , and INSERT operations.

- **void ackAsConsumed**

  To simplify the disaster recovery process of downstream SDK clients, the change tracking server supports consumption checkpoint storage for SDK clients. After a downstream SDK client encounters abnormal downtime and restarts, it automatically subscribes to and consumes data from the last consumption checkpoint that is recorded before downtime occurred.

  After message consumption is complete, you must call this method to send an ACK packet to instruct the DTS server to update the consumption checkpoints for the downstream SDK client. This ensures the integrity of the consumed data after an abnormal SDK client restarts.

Methods of the Record class

A Record object indicates a log entry in the RDS binlog file, such as BEGIN, COMMIT , and UPDATE operations.

- **String getAttribute(String key)**

  Obtains the main attribute values in a Record object. If the input parameter is an attribute name, the value of this attribute is returned.

  Table 13-11: Attribute names **describes the attributes that you can obtain by calling this method.**

  Table 13-11: Attribute names

Key	Description
record_id	The record ID. The ID does not ascend during the change tracking process.
instance	The database endpoint of the record. The format is <IP address>:<Port number>.
source_type	The database engine type of the record. Valid value: mysql.
source_category	The record type. Valid value: full_recorded.
timestamp	The time the record was written to the binlog. It is also the time the SQL statement was run in RDS.

Key	Description
checkpoint	The binlog file checkpoint of the record. The format is file_offset@file_name. The file_name parameter indicates the numeric suffix of the binlog file.
record_type	The operation type of the record. Valid values: insert, update, delete, replace, ddl, begin, commit, and heartbeat.
db	The database name of the table that is updated in the record.
table_name	The name of the table that is updated in the record.
record_recording	The encoding of the record.
primary	The primary key column value of the table that is updated in the record.
fields_enc	The encoded field values in the record. The fields are separated with commas (,). The value of a non-character field is null.

- Type getOpt()

  Obtains the operation type of the record, Valid values: insert, delete, update, replace, ddl, begin, commit, and heartbeat.

  The heartbeat is an exclusively defined indicator to reflect the health status of a change tracking channel. A heartbeat record is generated each second.
- String getCheckpoint()

  Obtains the checkpoint of the change record in the binlog file. The format of the returned checkpoint is binlog_offset@binlog_fid.

  binlog_offset indicates the offset of the change record in the binlog file, and binlog_fid indicates the numeric suffix of the binlog file. For example, if the binlog file name is mysql-bin.0008, the value of binlog_fid is 8.
- String gettimestamp()

  Obtains the timestamp of the change record in the binlog file.
- String getDbname()

  Obtains the database name of the table modified in the change record.
- String getTablename()

  Obtains the name of the table modified in the change record.

- **String getPrimaryKeys()**

  Obtains the primary key column value of the data entry that is modified in the change record. If the primary key is a composite key, the primary key column values are separated with commas (,).
- **DBType getDbType()**

  Obtains the database type of the change tracking instance. The value is MySQL because DTS only supports ApsaraDB RDS for MySQL.
- **String getServerId()**

  Obtains the IP address and port number that the ApsaraDB RDS for MySQL instance uses to run the process corresponding to the change record. The format is <IP address>:<Port number>.
- **int getFieldCount()**

  Obtains the number of fields that are changed in the record.
- **List<Field> getFieldList()**

  Obtains a list of fields. The data type of the returned value is List<Field>.

  List<Field> contains the definitions of all fields that are changed in the record and the image values before and after the change. For more information about the Field class, see *Methods of the Field class*.
- **Boolean isFirstInLogevent()**

  Checks whether the record is the first transaction log entry in a large volume of database changes. If yes, True is returned. If no, False is returned.

Methods of the Field class

The Field class defines the attributes of each field, such as the code, type, name, value, and whether the field is a primary key field. This section defines the methods of the Field class.

- **String getEncoding()**

  Obtains the encoding format of the field value.
- **String getFieldname()**

  Obtains the name of this field.
- **Type getType()**

  Obtains the data type of the field.

- **ByteString getValue()**

  Obtains the value of this field. The return type is ByteString. If the field is not specified, NULL is returned.

- **Boolean isPrimary()**

  Checks whether the field is a primary key field of the table. If yes, True is returned. If no, False is returned.

13.5.6 SDK quick start

This section describes how to use the DTS Java SDK to perform some basic operations.

Initialize RegionContext

RegionContext is used to save and set security authentication information and the network access mode. The following code is used to initialize RegionContext to set the security authentication credentials and network access mode.

```
import java.util.List;
import com.aliyun.drc.clusterclient.ClusterClient;
import com.aliyun.drc.clusterclient.DefaultClusterClient;
import com.aliyun.drc.clusterclient.RegionContext;
public class MainClass
{
      public static void main(String[] args) throws Exception {
        // Create a RegionContext.
          RegionContext context = new RegionContext();
          context.setAccessKey("<AccessKey>");
          context.setSecret("<AccessKeySecret>");
          context.setUsePublicIp(true);
          // Create a subscription client.
          final ClusterClient client = new DefaultClusterClient(
context);
          // The following is other invocation code:
          ...
    }
}
```

Initialize Listeners

The functions of data consumption are implemented through the Listener class. After ClusterClient is initialized, add a Listener class, which defines the notify function to receive and consume subscribed data. The following code demonstrates the most basic consumption logic, and is used to display the subscribed data on the screen.

```
import com.aliyun.drc.clusterclient.ClusterClient;
import com.aliyun.drc.clusterclient.ClusterListener;
import com.aliyun.drc.clusterclient.DefaultClusterClient;
```

```
import com.aliyun.drc.clusterclient.RegionContext;
import com.aliyun.drc.clusterclient.message.ClusterMessage;
public class MainClass
{
    public static void main(String[] args) throws Exception {
        // Initialize a RegionContext object
        ...
        //Initialize a ClusterClient object
        ...
        ClusterListener listener = new ClusterListener(){
             @Override
             public void notify(List<ClusterMessage> messages) throws
Exception {
                  for (ClusterMessage message : messages) {
                    // Display the subscribed incremental data.
                      System.out.println(message.getRecord() + ":" +
message.getRecord().getTablename() + ":"
                      + message.getRecord().getOpt());
                      // After data consumption is completed, send an
ACK packet to DTS by calling
                      message.ackAsConsumed();
               }
      }
     }
}
```

DTS saves the consumption checkpoints of the SDK to the DTS server. This simplifies disaster recovery during the use of the SDK. The askAsConsumed() interface in the preceding sample code reports the consumption checkpoint of the last data record consumed by the SDK to the DTS server. When the SDK restarts after an unexpected downtime, it automatically obtains the consumptio n checkpoint from the DTS server and restarts from this checkpoint to avoid data duplication.

Start ClusterClient

Use the following code:

```
import java.util.List;
import com.aliyun.drc.clusterclient.ClusterClient;
import com.aliyun.drc.clusterclient.ClusterListener;
import com.aliyun.drc.clusterclient.DefaultClusterClient;
import com.aliyun.drc.clusterclient.RegionContext;
import com.aliyun.drc.clusterclient.message.ClusterMessage;
import org.slf4j.Logger;
import org.slf4j.LoggerFactory;
public class MainClass
{
   public static void main(String[] args) throws Exception {
    // Initialize RegionContext.
    ...
    // Initialize ClusterClient.
    ...
    // Initialize ClusterListener.
    ...
    // Add listeners.
      client.addConcurrentListener(listener);
```

```
      // Set the requested subscription channel ID.
      client.askForGUID("dts_rdsrjiei2u2afnb_DSF");
      // Start a background thread. Note that there will be no
blocking and the main thread cannot exit.
      client.start();
}
```

In the preceding code, the askForGUID() interface sets the subscription channel ID requested by the client. The ID of this subscription channel is obtained from the DTS console. Once the subscription channel ID is configured, the SDK can obtain the incremental data through the subscription channel.

You must add a Listener class to a client before starting the client. In this way, when the client pulls incremental data from the subscription channel, it starts data consumption by calling the notify method of the Listener synchronously.

13.5.7 Use SDK to track data changes

You can use the SDK to track data changes. DTS records the tracked data changes in a custom format. This topic describes how to parse various types of SQL statements.

Parse a DDL statement

If a record is a DDL statement, the operation type of this record is DDL. The DDL statement is stored in the value of the first column. You can use the following sample code to obtain the DDL statement:

```
String ddl_string;
Record.Type type=record.getOpt();
if(type.equals(Record.Type.DDL)){
    List<DataMessage.Record.Field> fields = record.getFieldList();
    ddl_string = fields.get(0).getValue().toString();
}
```

Parse an INSERT statement

If a record is an INSERT statement, the operation type of this record is INSERT. You can use the following sample code to obtain the complete INSERT statement:

```
StringBuilder insert_string=new StringBuilder();
Record.Type type=record.getOpt();
DataMessage.Record.Field field;
StringBuilder FieldName=new StringBuilder();
StringBuilder FieldValue = new StringBuilder();
if(type.equals(Record.Type.INSERT)){
       int i=0;
       List<DataMessage.Record.Field> fields = record.getFieldList();

       for (; i < fields.size(); i++) {
              field = fields.get(i);
  FieldName.append('`'+field.getFieldname().toLowerCase()+'`');
              FieldValue.append(field.getValue());
```

```
                if (i ! = fields.size() - 1) {
                        FieldName.append(',');
                        FieldValue.append(',');
                }
        }
        insert_string.append("insert "+ record.getTablename()+"("+
FieldName.toString()+") values("+FieldValue.toString()+");");
}
```

Parse an UPDATE statement

If a record is an UPDATE statement, the operation type of this record is UPDATE.

The field values prior to the UPDATE operation are stored in Record.getFieldList() entries with even indexes. The field values after the UPDATE operation are stored in Record.getFieldList() entries with odd indexes.

You can use the following sample code to obtain the complete UPDATE statement if the updated table has a primary key:

```
StringBuilder update_string=new StringBuilder();
Record.Type type=record.getOpt();
DataMessage.Record.Field field;
StringBuilder SetValue = new StringBuilder();
StringBuilder WhereCondition = new StringBuilder();
String ConditionStr;
boolean hasPk=false;
boolean pkMode=false;
boolean hasSet=false;
if(type.equals(Record.Type.UPDATE)){
    int i=0;
    DataMessage.Record.Field OldField = null;
    DataMessage.Record.Field NewField = null;
    List<DataMessage.Record.Field> fields = record.getFieldList();
    for (; i <fields.size() ; i++) {
        if (i % 2 == 0) {
            OldField = fields.get(i);
            continue;
        }
    NewField = fields.get(i);
    if (field.isPrimary()) {
        if (hasPk) {
            WhereCondition.append(" and ");
        }
        //where old value
        ConditionStr = getFieldValue(OldField);
        if(ConditionStr==null){
  WhereCondition.append("`"+field.getFieldname().toLowerCase()+"`" + "
 " + "is null");
        }else{
               WhereCondition.append("`"+field.getFieldname().
toLowerCase()+"`"+" = "+ NewField.getValue());
         }
        hasPk = true;
    }
    if (hasSet) {
        SetValue.append(COMMA);
    }
```

```
      SetValue.append("`"+field.getFieldname().toLowerCase()+"`" + " =
 " + field.getValue());
    String setStr = getFieldValue(field);
    hasSet = true;
    }
    update_string.append("Update "+record.getTablename() +" Set " +
SetValue + " Where "+WhereCondition +";");
}
```

Parse a DELETE statement

If a record is a DELETE statement, the operation type of this record is DELETE. You can use the following sample code to obtain the complete DELETE statement if the deleted table has a primary key:

```
StringBuilder delete_string=new StringBuilder();
Record.Type type=record.getOpt();
DataMessage.Record.Field field;
StringBuilder FieldName=new StringBuilder();
StringBuilder FieldValue = new StringBuilder();
StringBuilder DeleteCondition = new StringBuilder();
boolean hasPk=false;
boolean pkMode=false;
if(type.equals(Record.Type.DELETE)){
   int i=0;
   List<DataMessage.Record.Field> fields = record.getFieldList();

   delete_string.append("Delete From" + record.getTablename() + "where
");
   // Check whether the table has a primary key.
   if (record.getPrimaryKeys() ! = null) {
             pkMode = record.getPrimaryKeys().length() > 0 ? true :
false;
   }
   for (; i < fields.size(); i++) {
            if ((pkMode && ! field.isPrimary())) {
                    continue;
            }
            if (hasPk) {
                    delete_string.append(" and ");
            }
            delete_string.append(field.getFieldname() + "=" + field.
getValue());
            hasPk = true;
    }
    delete_string.append(";");
}
```

Parse a REPLACE statement

If a REPLACE statement has been executed for the source database, the operation type of this record is UPDATE or INSERT. If the value specified in the REPLACE statement does not exist, the record operation type is INSERT. If the value specified in the REPLACE statement already exists, the record operation type is UPDATE.

Parse a BEGIN statement

If a record is a BEGIN statement, the operation type of this record is BEGIN. You do not need to perform any operations on fields because the BEGIN statement does not modify fields. You only need to determine that the operation is a BEGIN operation. You can use the following sample code to obtain the BEGIN statement:

```
StringBuilder sql_string = new StringBuilder();
Record.Type type = record.getOpt();
if(type.equals(Record.Type.BEGIN)){
        sql_string.append("Begin");
}
```

Parse a COMMIT statement

If a record is a COMMIT statement, the operation type of this record is COMMIT. You do not need to perform any operations on fields because the COMMIT statement does not modify fields. You only need to determine that the operation is a COMMIT operation. You can use the following sample code to obtain the COMMIT statement:

```
StringBuilder sql_string = new StringBuilder();
Record.Type type = record.getOpt();
if(type.equals(Record.Type.COMMIT)){
        sql_string.append("commit");
}
```

13.5.8 Run the SDK demo code

This section describes how to run the demo code provided by the DTS console.

1. **Create an AccessKey.**

   Your account must pass the AccessKey authentication before you can use an SDK to connect to a subscription channel. Therefore, before using the SDK, you must obtain an AccessKey. For more information, see the "Obtain an AccessKey" section of the *DTS Developer Guide* **.**

2. **Install the Java SDK.**

   The development environment supported by the DTS Java SDK is J2SE Development Kit (JDK) V1.5 or later.

   For an Eclipse project, you can follow these steps to install the Java SDK:

   a. **Click View Example Code and download the SDK package** *consumer.jar*.
   b. **Import the JAR package to an Eclipse project as follows:**

      In Eclipse, right-click your project and choose Properties > Java Build Path > Libraries > Add External JARs. Select the path for storing the *consumer.jar* **package** consumer.jar.
   c. **Select the** *consumer.jar* **package and click OK.**

      Then you can use the DTS Java SDK in the project.

3. **Run the demo code.**

   DTS provides the SDK demo code. You can copy the demo code by using the View Demo Code option in the DTS console. For an Eclipse project, you can follow these steps to run the demo code:

   a. **Create a class named MainClass in the src directory of the Eclipse project.**
   b. **Open the generated Java file** *MainClass* **and delete the code template.**
   c. **Paste the demo code into the** *MainClass* **file.**
   d. **Modify the AccessKeyId, AccessKeySecret, and subscription channel ID in the demo code.**

      Change the marked parts in the preceding demo code to the AccessKeyId, AccessKeySecret, and subscription channel ID of your account.

      You can obtain the subscription channel ID from the *DTS console*.
   e. **In Eclipse, right-click the demo file and choose Run as > Java Application to run the demo code.**

14 Data Management Service (DMS)

14.1 What is Data Management Service?

Data Management Service (DMS) is an integrated database solution that includes data, schema, and server management, access control, BI insights, data trend analysis, data tracking, and performance optimization. DMS can be used to manage relational databases and NoSQL databases, such as MySQL, PostgreSQL, MongoDB, and Redis. DMS can also be used to manage Linux servers.

DMS console

Figure 14-1: DMS console for relational databases **shows the DMS console for relational databases.**

Figure 14-1: DMS console for relational databases

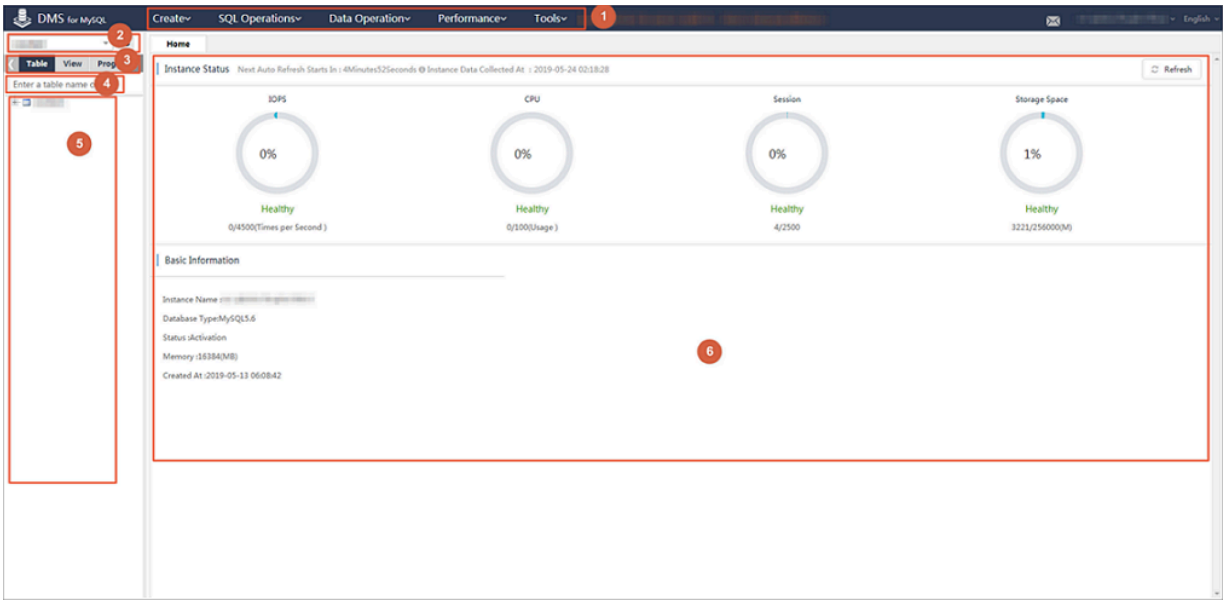


Table 14-1: Description of functional modules **describes the functional modules.**

Table 14-1: Description of functional modules

No.	Name	Description
1	Top navigation bar	Provides an entry to the functional modules of DMS.

No.	Name	Description
2	Database drop-down list	Allows you to select a database from the list to access the tables and data objects in the database.
3	Navigation buttons for database objects	Allows you to navigate to tables, views , and programmable objects such as functions, stored procedures, triggers, and events.
4	Table search box	Provides fuzzy matching to quickly locate tables.
5	DMS object list	Displays the details of database objects such as tables.
6	Instance health report	Displays the current health status of the RDS instance.

Supported database types

- **MySQL**
- **PostgreSQL and PPAS**

Supported database operations

- **SQL operations**
  - **Use of SQL editor**
  - **Use of SQL command line interface**
  - **Saving of work environment settings**
  - **SQL execution**
  - **SQL optimization**
  - **SQL formatting (SQL statement improvement)**
  - **Viewing of execution plans**
  - **Smart SQL completion**
- **Operations on database objects**
  - **Operations on data tables**
  - **Operations on table schemas: creation and deletion of tables and modification of table schemas**
  - **Changes to table data: insertion, update, and deletion of data**
  - **Table data query and visualized editing**

- **Operations on views and programmable objects such as functions, stored procedures, triggers, and events**
  - **Creation**
  - **Modification**
  - **Deletion**
  - **Enabling and disabling**
- **Data processing**
  - **Data import**
  - **Data export**
- **Performance and diagnostics**
  - **Real-time performance**
  - **Real-time session**
  - **Lock wait analysis**
- **Use of data processing tools**
  - **Drawing of ER diagrams**
  - **Collection of statistics on table data volumes**
  - **Batch operations on tables**

User-friendly interaction

DMS provides user-friendly tips. When an error occurs, DMS displays suggestions to guide you to complete your goal.

14.2 Log on to an RDS instance through DMS

This topic describes how to log on to an RDS instance through DMS.

Context

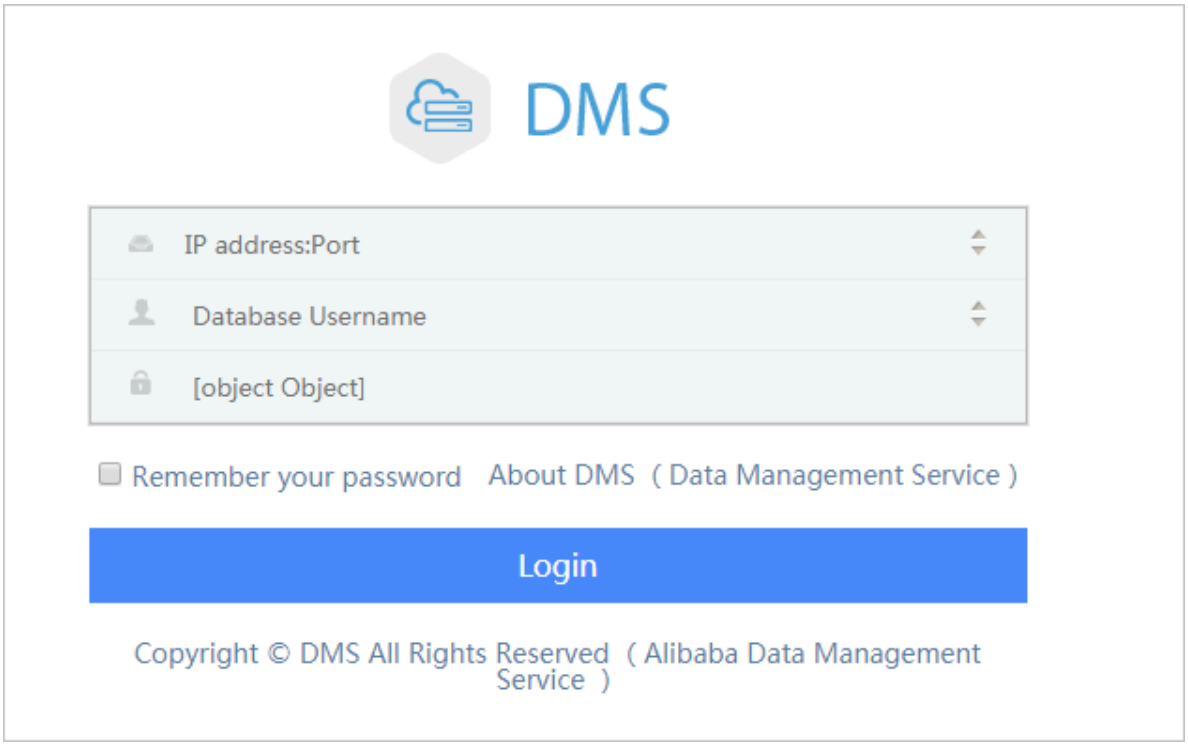
The RDS console provides the option to log on to RDS instances by using DMS. You can use DMS to manage database data and schemas. It can manage relational databases such as MySQL, and PostgreSQL, and OLAP databases.

Procedure

1. **Log on to the RDS console. For more information, see "Log on to the RDS console" in the** *RDS User Guide* **.**

2. **Click the ID of the target instance or click the management icon in the Actions column corresponding to the instance, and choose View Details from the shortcut menu. The Basic Information page appears.**
3. **Click Log On to DMS to go to the logon page of the DMS console.**
4. **Enter the logon information.**

Figure 14-2: DMS logon page



The following table describes the parameters on the DMS logon page.

No.	Description
1	**The internal or external network address and corresponding port number of the target RDS instance, such as rm-test00000k012.mysql.aliyun-inc.com:3306. To obtain the internal or external network address and its port number, perform the following steps:** **a. Log on to the RDS console.** **b. Click the ID of the target instance or click the management icon in the Actions column corresponding to the instance, and choose View Details from the shortcut menu. The Basic Information page appears.** **c. Query the network address and port number in the Internal Network Connection Information section.**

No.	Description
2	The account that is used to connect to the instance. **Note:** The account is created in the RDS instance. For information about how to create an account for a MySQL instance, see "Create a standard account" in the *RDS User Guide* .
3	The password of the account. **Note:** The password is specified when you create the account in the RDS instance.
4	The type of the target database.

5. Click Log On.

Note:
If you want the browser to remember the password, select `Remember Password`, and then click Log On.

14.3 SQL operations

14.3.1 Use the command window

This topic briefly describes how to use the DMS command window.

Context

A MySQL database is used as an example.

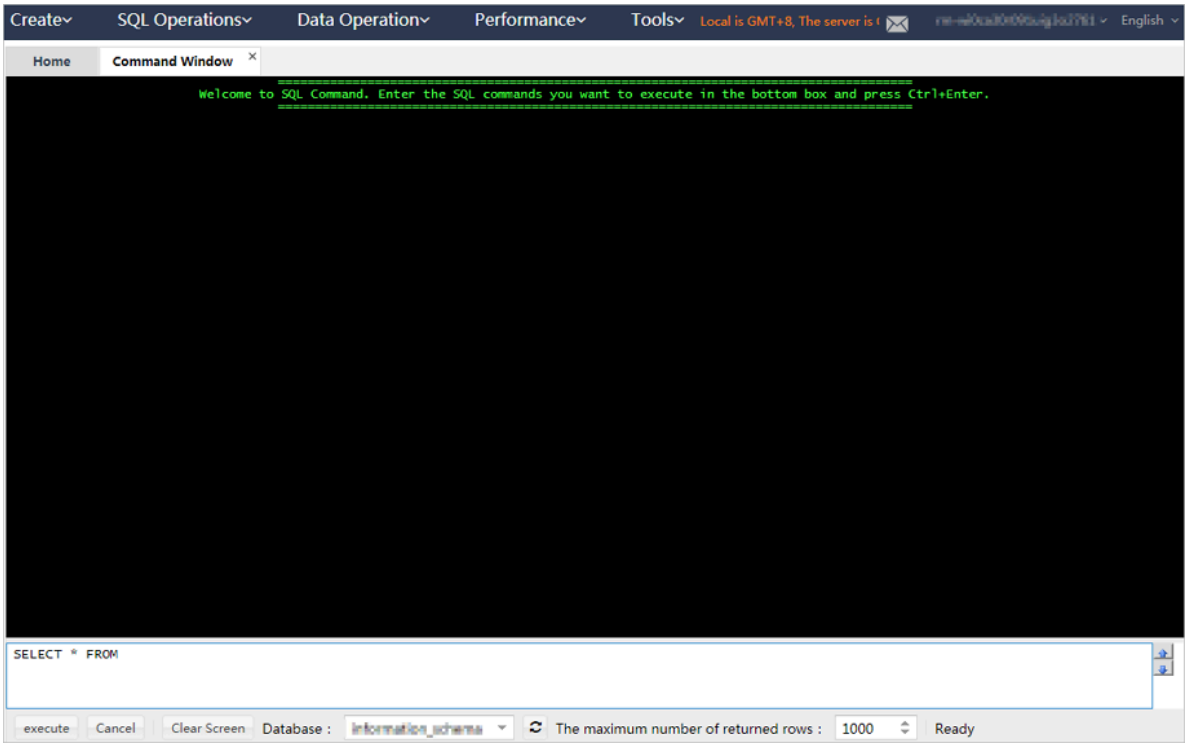
Procedure

1. *Log on to an RDS instance through DMS*.

2. **In the top navigation bar, choose SQL Operations > Command Window.**

   An empty command window appears, as shown in *Figure 14-3: Command window*.

   Figure 14-3: Command window



3. **Enter an SQL statement in the command window, click Execute, and view the execution result, as shown in** *Figure 14-4: Execution of a SQL statement*.

   Figure 14-4: Execution of a SQL statement

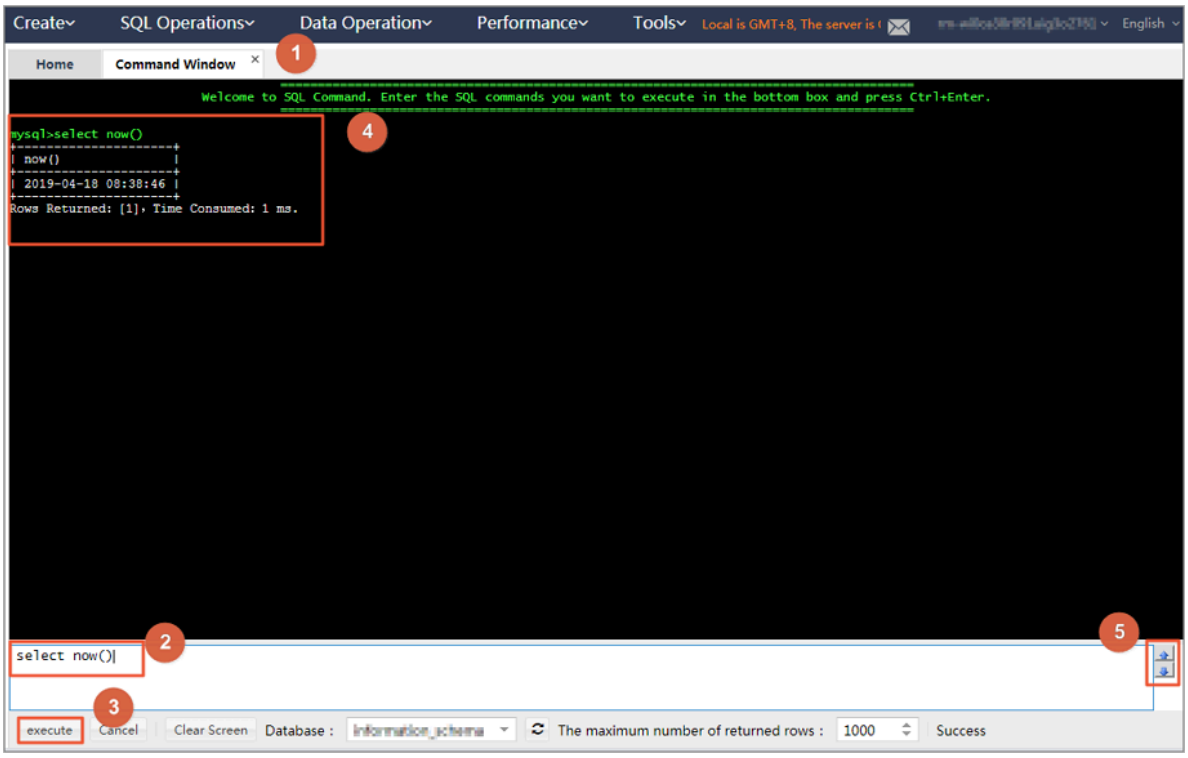


Table 14-2: Description of the numbered items **describes the numbered items.**

Table 14-2: Description of the numbered items

No.	Name	Description
1	Command window	Displays the execution results of SQL statements.
2	SQL statement input area	Offers an area to enter SQL statements.
3	`Execute` button	Executes the entered SQL statements.
4	Result display area	Adds execution results to `Result Area`.
5	`Up` and `Down` arrows	You can click the up or down arrow to view an executed SQL statement and execute it again.

4. **Optional: If the execution process takes longer time than expected, you can click Cancel to abort the execution.**

5. **Optional: Click Clear Screen to clear the results for proper display of subsequent results.**

   To switch to another database, select the new database from the `Database` drop-down list.

14.3.2 Use the SQL window

14.3.2.1 Open an empty SQL window

This topic describes how to use SQL windows.

Context

- **A MySQL example is used as an example.**
- **A maximum of 20 SQL windows (including the homepage) can be opened at the same time in DMS. We recommend that you open no more than five SQL windows .**

Procedure

1. *Log on to an RDS instance through DMS*.

2. **In the top navigation bar, choose SQL Operations > SQL Window to open an SQL window.**

   Figure 14-5: Empty SQL window **shows the empty** `SQL window` **you opened.**

   Figure 14-5: Empty SQL window

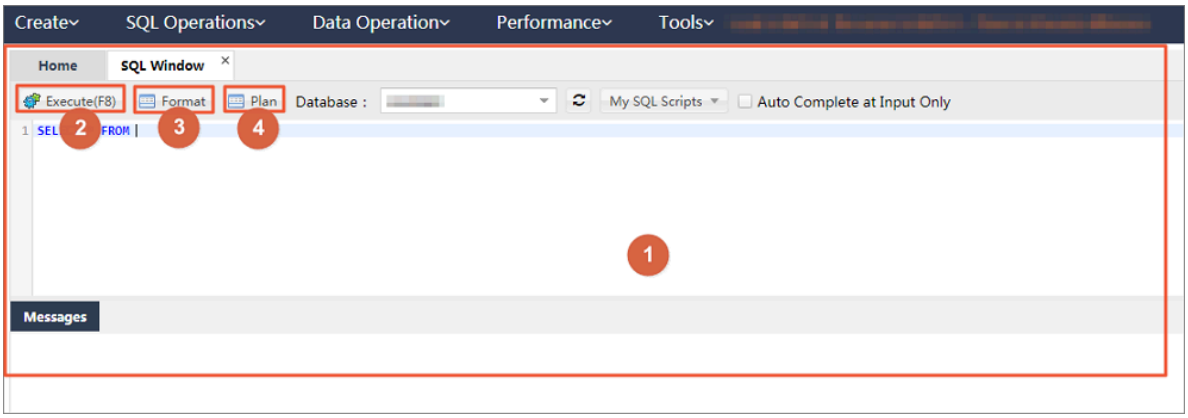


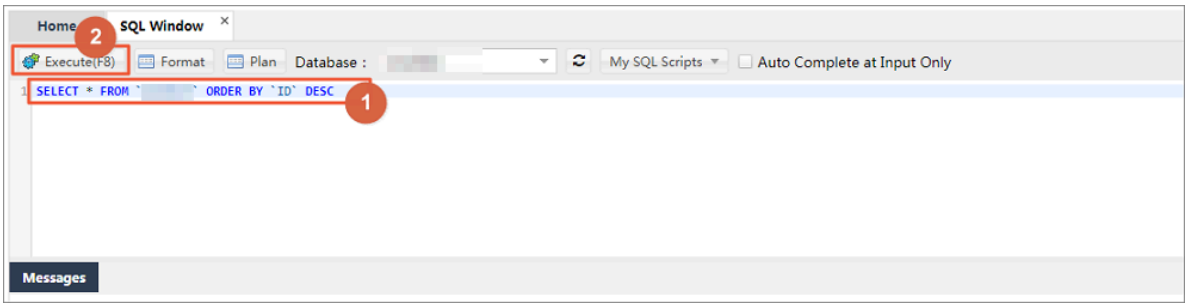
   Table 14-3: Numbered items in the SQL window **describes the numbered items in the SQL window.**

   Table 14-3: Numbered items in the SQL window

No.	Name	Description
1	SQL window	The green-framed area is the main body of the SQL window.
2	Run (F8) button	Click this button to run the entered SQL statements.
3	Format button	Click this button to format the entered SQL statements to make them more readable.
4	Execution Plan button	Click this button to display the execution plans of the selected SQL statements. You can optimize the SQL statements and improve SQL processing performance based on the execution plans.

3. **Enter the SQL statement you want to execute and click Run to complete the SQL query or update, as shown in *Figure 14-6: Run the SQL statement*.**

   Figure 14-6: Run the SQL statement



4. **You can view the result set, as shown in *Figure 14-7: View the result set*.**

   Figure 14-7: View the result set

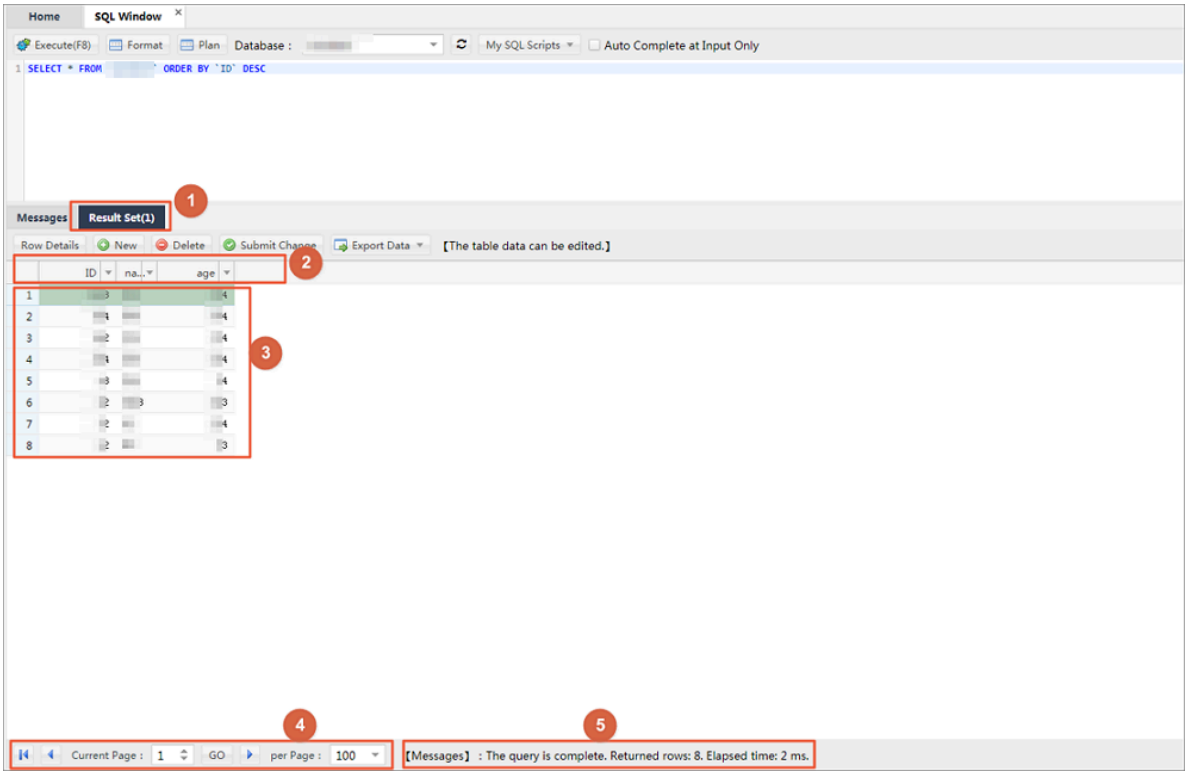


Table 14-4: Description of the numbered items in the result set

No.	Description
1	The Result Set tab shows the results returned by the SQL query statement.

No.	Description
2	The first row of the table shows the field names. If an alias has been specified for a field in the SQL statement, the alias is displayed in this table.
3	The data area of the table shows the query results row by row . If the data area is not big enough to show the full results, horizontal and vertical scroll bars will appear to help you navigate the results.
4	Click Show in Pages or Next Page to view the results. · Each page shows 100 query results by default. Go to the next page to view more results. · You can set the number of results displayed per page as needed. · The results on the next page are appended to the table numbered 3 in the figure.
5	Progress of result acquisition and time elapsed.

5. **View the message about SQL execution.**

   Each time a data query (SELECT) or data correction (INSERT, UPDATE, or DELETE) statement is executed, DMS returns a message about the execution, including the status and impact.

   Figure 14-8: Data query **shows the message returned for data query.**

   Figure 14-8: Data query

   Figure 14-9: Data correction **shows the message returned for data correction.**

   Figure 14-9: Data correction

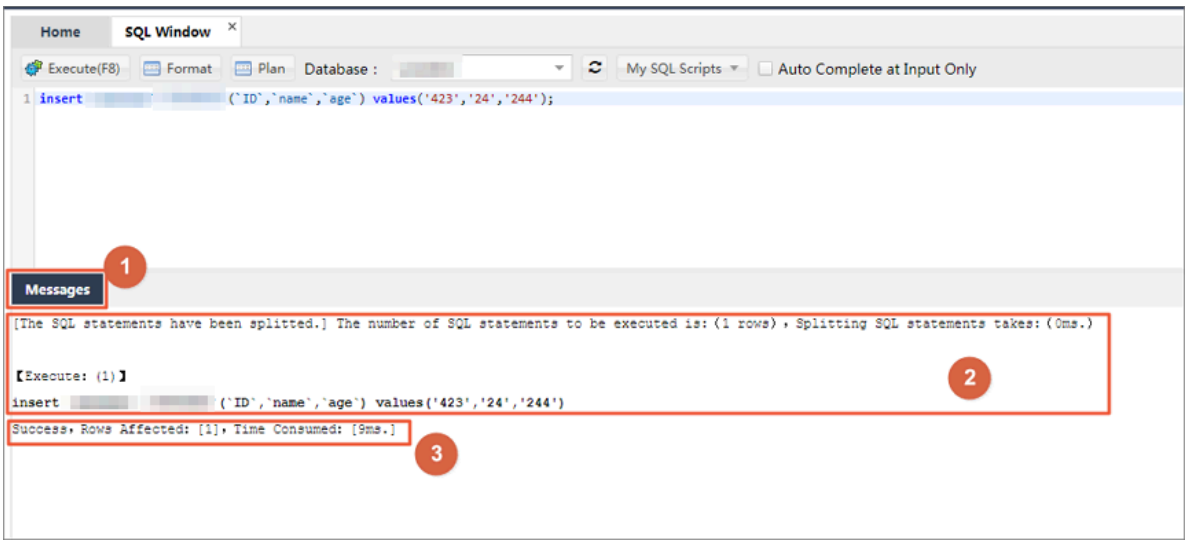


   Table 14-5: Description of the numbered items in the data correction window **describes the numbered items in the data correction window.**

Table 14-5: Description of the numbered items in the data correction window

No.	Description
1	After you run an SQL statement, you can click the Message tab to view the execution status. No result set is returned for data correction. DMS displays a message after data correction is complete.
2	DMS runs the entered SQL statements step by step. · Analyzes the entered SQL statements. · Runs the SQL statements in the database. · Displays the queried data. · Displays statistics. For example, the number of data rows that are queried or affected.
3	DMS displays the SQL execution results. · Whether the execution is successful. · Number of queried rows, or number of rows affected by the Add, Delete, or Modify operation. · Time consumed to run the SQL statements.

6. **Run SQL statements in batches.**

   DMS supports batch execution of SQL statements, as shown in *Figure 14-10: Batch execution*.

   Figure 14-10: Batch execution

   - **1: Shows the execution results of the first SQL statement.**
   - **2: Shows the execution results of the second SQL statement.**

   a) **Separate each SQL statement with a semicolon (;) or another separator.**

   b) **If you want to run only some SQL statements, select the SQL statements you want to run. If you want to run all SQL statements, deselect or select all SQL statements, and click Run.**

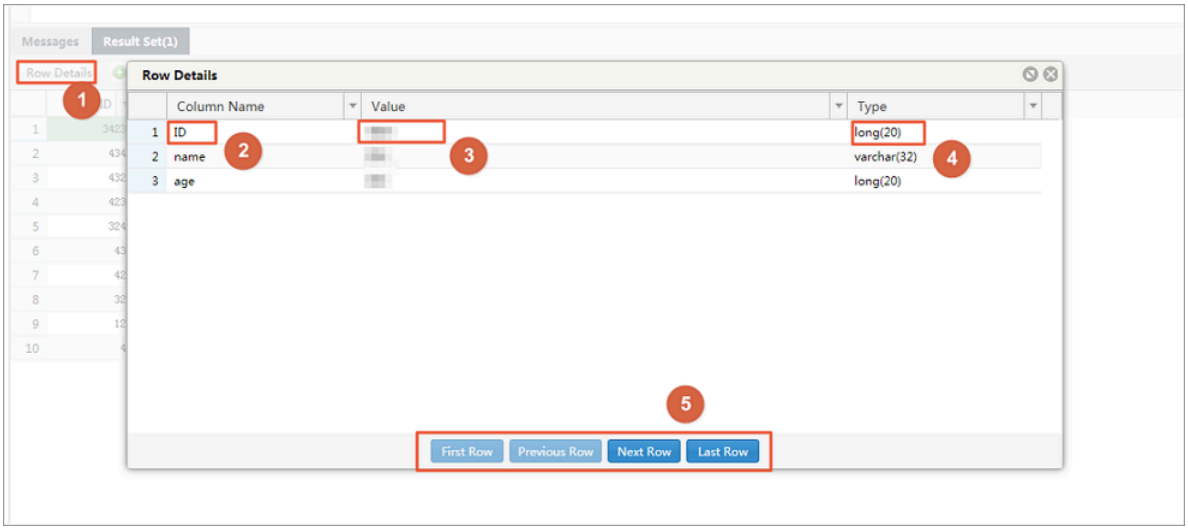
      Wait until all SQL statements are executed.

   c) **View the execution results.**

      If you run the SELECT statement, DMS displays the result set. If you run other statements, DMS displays the execution results, such as the number of affected rows.

7. **Click Single Row Details to view the details of a single record in the result set, as shown in** *Figure 14-11: Single row details*.

Figure 14-11: Single row details



The following table describes the numbered items in the Single Row Details window.

Table 14-6: Description of the numbered items in the Single Row Details window

No.	Description
1	Select the single row record you want to display in the `Result Set` table, and click Single Row Details to view a single data record. The Single Row Details dialog box displays every `Field name`, `Field value`, and `Field type` of the record.
2	`Field name`: If you have specified aliases for fields, the aliases are displayed.
3	`Field value`: DMS automatically parses and displays the field values. Data such as time and binary code is formatted as a string for clear display.
4	`Field type`: You can view the type and length of each field.
5	Record navigation area. The Previous, Next, First, and Last buttons make it easier for you to view single row details of previous and subsequent records.

8. **Optional: Edit the queried data in the result set.**

   - **Click Add to add a row of data to the currently queried table.**
   - **Click Delete to delete the selected row of data from the result-set table.**
   - **Select the row that you want to perform operations on.**
   - **Update the field values in the selected row directly.**

   After you modify data, click Submit Changes to save the changed results to the database.

   After you click Submit Changes, DMS displays the SQL statements required to save your changes. This allows you to confirm the changes and prevent misoperations that cause loss of data.

   Click OK to apply the changes to the database as expected.

9. **Click Beautify to improve the readability and writability of the selected SQL statements.**

   - **Only the selected SQL statements are beautified. If you do not select any SQL statements, all the SQL statements that you entered are beautified.**
   - **The beautify function reformats your SQL statements to standard and readable statements, without changing the SQL execution logic and semantics or affecting the execution.**

   Examples:

   Figure 14-12: Original SQL statements **shows the original SQL statements.**

   Figure 14-12: Original SQL statements

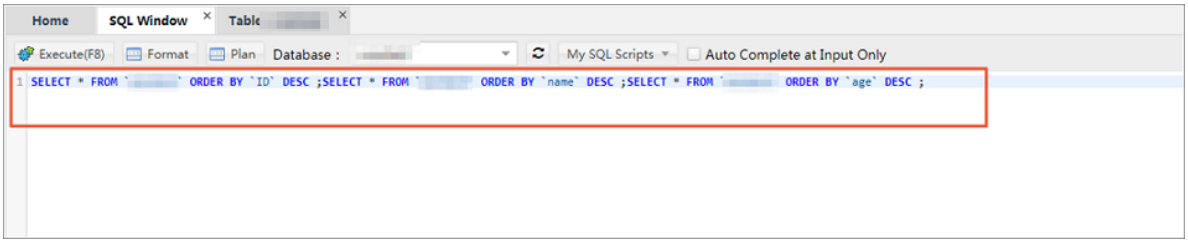
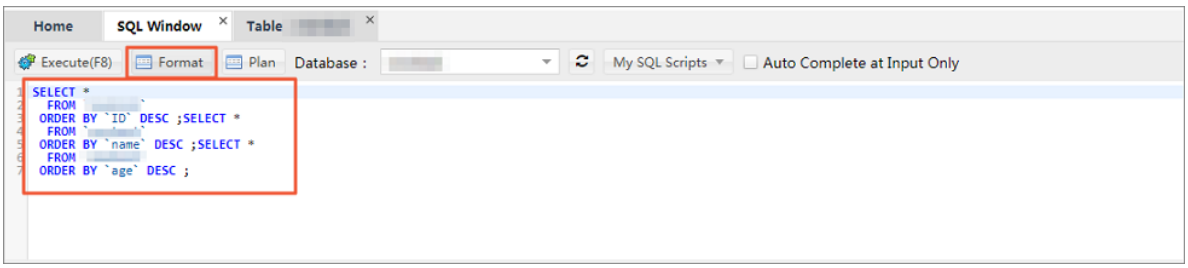


   Figure 14-13: Beautified SQL statements **shows the beautified SQL statements.**

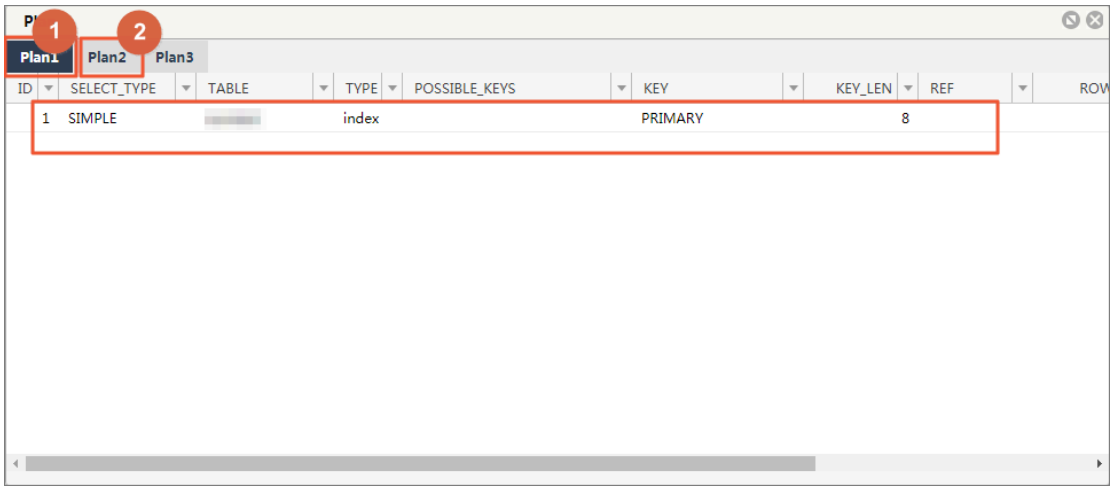
   Figure 14-13: Beautified SQL statements



10. Click Plan to view the execution plan if you want to troubleshoot SQL-related problems or optimize SQL performance.

- After you click Plan, DMS displays the execution plan of the selected SQL statement. If no SQL statement is selected, DMS displays the execution plans of all SQL statements.
- DMS displays an execution plan in detail. You can view information such as the type of the execution plan and possible keys.
  - Different databases display execution plans in different ways and to varying extents.
  - If you want to view the execution plans of several SQL statements, DMS displays the execution plan of each SQL statement in detail on different tabs, as shown in *Figure 14-14: Execution plan*.

Figure 14-14: Execution plan



■ 1: Shows the execution plan of the first SQL statement in detail.

■ 2: Shows the execution plan of the second SQL statement in detail.

14.3.2.2 Restore a saved SQL window

This topic describes how to restore a saved SQL window.

Context

- A MySQL example is used as an example.
- A maximum of 20 SQL windows (including the homepage) can be opened at the same time in DMS. We recommend that you open no more than five SQL windows .

Procedure

1. *Log on to an RDS instance through DMS*.
2. In the top navigation bar, choose SQL Operations > SQL Window.
3. Save the operating environment of the current SQL window.
   - DMS automatically saves the work environment when you close the operation page.
   - When you log on to the DMS console next time, DMS automatically restores the last work environment, including the last used database, the SQL windows you opened, and the SQL statements you entered in the SQL windows.
   - When you close a SQL window, DMS prompts you to confirm whether you want to save the content within the window.
     - 1: Click the Close icon in the upper-right corner of the SQL window to close the window.
     - 2: DMS prompts you to confirm whether you want to save the work content. Click Close and Save. DMS then saves the work content of the SQL window, and closes the window after the work content is saved.

       If you click Close, DMS does not save the present work in the SQL window.
4. Restore the saved SQL window.
   a) Choose SQL Operations > Saved SQL Windows.

      DMS displays all the saved SQL windows.
   b) Click New SQL Window to restore one of the saved SQL windows.
   c) When you log on to the database through DMS, DMS automatically restores the work content of the last saved SQL window.

14.3.2.3 Manage frequently used SQL commands

This topic describes how to manage frequently used SQL commands in DMS.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an RDS instance through DMS*.
2. In the top navigation bar, choose SQL Operations > SQL Window to open an SQL window.

3. **Perform the following operations:**

   - **Add a frequently used SQL command.**

     Choose My SQL > Add My SQL to add a frequently used SQL command.

     - **Applicable scope: The custom SQL command is applicable to all scenarios.**
     - **All databases: You can access the custom SQL command in any databases that you log on to from DMS.**
     - **Current instance: You can access the custom SQL command only through the currently connected instance (with an IP address and a port number).**
     - **Current database: You can access the custom SQL command only through the currently connected database. If you switch to another database, choose My SQL > Select My SQL. The custom SQL command is not displayed.**

   - **View saved SQL commands.**

     Choose My SQL > Select My SQL to view the frequently used SQL commands you saved.

   - **Manage your SQL commands.**

     Choose My SQL > Manage My SQL to manage frequently used SQL commands.

     - **On the Manage My SQL page, click Edit or Delete to edit or delete your SQL commands.**
     - **On the Manage My SQL page, click Add to add an SQL command.**
     - **Double-click an SQL command under `My SQL` to insert the command into the SQL Window. The command is in the selected state in the SQL window.**

14.3.2.4 Use the SQL template

This topic describes how to use the SQL template in DMS.

Procedure

1. *Log on to an RDS instance through DMS*.
2. **In the top navigation bar, choose SQL Operations > SQL Window. A SQL window appears.**

   The SQL template is displayed in the rightmost part of the SQL window.

3. **Double-click an SQL command or drag it into the SQL window. Then you can use or reference the command.**

   You can directly modify the commands referenced from the template even if you are not familiar with the commands.

14.3.3 Table operations (based on the Table directory tree)

14.3.3.1 Open a table-based SQL window

This topic describes how to open a table-based SQL window in DMS.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an RDS instance through DMS*.
2. **In the left-side directory tree, right-click a table and choose SQL Operation Data from the shortcut menu to open an SQL window.**

   DMS automatically runs the SQL statement that queries top 50 records of the table.

14.3.3.2 Edit table data

This topic describes how to manage frequently used SQL commands in DMS.

Context

- **A MySQL database is used as an example.**
- **This function applies to tables with average data volumes. For tables containing large volume of data, locate the data before editing. Data locating may take some time.**

Procedure

1. *Log on to an RDS instance through DMS*.

2. **In the left-side directory tree, right-click a table and choose Open Table from the shortcut menu.**

   A window appears, indicating the data of the selected table.

   - **1: In the left-side directory tree, right-click a table and choose Open Table from the shortcut menu. The data edit window appears.**
   - **2: You can modify the values of the fields in the table.**
   - **3: After you modify the data, click Submit Changes to submit the modified data.**

14.4 Database development

14.4.1 Overview

This topic describes how to add, modify, delete, and manage objects such as indexes, foreign keys, and stored procedures.

14.4.2 Table

14.4.2.1 Create a table

This topic describes how to create a table in DMS.

Procedure

1. *Log on to an RDS instance through DMS*.
2. **You can create a table through any of the following methods:**
   - **In the top navigation bar, choose Create > Table.**
   - **In the left-side Table directory tree, right-click a table and choose Add Table from the shortcut menu.**
   - **In the Common Operations area of the homepage, click Create Table.**
3. **Edit columns.**

   Go to the Create: Table page, which displays the Column Info tab by default.

   You can edit the basic information and extended information of the fields as needed.

   You can also click Column Info to edit the table information.

4. Click the Index tab to edit indexes.
   - Click Add to add an index.
   - Click Delete to delete an index.
   - You can edit the index row to modify index information.
5. Click the Foreign Key tab to go to the Edit Foreign Keys tab page.
   - Click Add to add a foreign key. The new key is editable.
   - Click Delete to delete a foreign key.
   - You can edit the index row to modify index information. When you edit a foreign key, enter the key name, column name, and information of the referenced databases, tables, and columns.
6. Click the Partition tab and enter the SQL information of the partition.
7. Click the Basic Info tab to edit the basic information of the table.
   - You can edit the table name, storage engine, character set, and description.
   - You can click More to edit table parameters.
8. Click Save. DMS generates the SQL statements used to create a table.

   Click OK after you confirm the SQL statements. DMS then adds the table to your database.

14.4.2.2 Edit a table

This topic describes how to edit a table in DMS.

Procedure

1. *Log on to an RDS instance through DMS*.
2. In the left-side Table directory tree, right-click a table and choose Edit Table Structure from the shortcut menu to edit the table structure.
3. The Edit Table window is similar to the Create Table window. DMS automatically loads the table structure into the window.
   - 1: Select a table object type, such as `Column Info` or `Index`.
   - 2: Click a specific operation on the table object, which is similar to the Create and Edit operations on tables.
   - 3: Click Open Table Data to view and modify table data.
   - 4: Click Create Statement to view the statements used to create a table.

4. Click Save. DMS displays the SQL statements used to modify the table structure.

   Click OK after you confirm the SQL statements. DMS then saves the modified table structure to your database.

14.4.2.3 Delete a table

This topic describes how to delete a table in DMS.

Procedure

1. *Log on to an RDS instance through DMS*.
2. In the left-side Table directory tree, right-click the table you want to delete and choose Delete Table from the shortcut menu.

   
   Warning:
   Deleting tables is a high-risk operation. Therefore, exercise caution when deleting tables.
3. Click Yes to delete the table.

14.4.2.4 Create a similar table

This topic describes how to duplicate a table in DMS.

Procedure

1. *Log on to an RDS instance through DMS*.
2. In the left-side Table directory tree, right-click the table you want to copy and choose Create Table Like from the shortcut menu.

   The Create Similar Table window appears.
3. Enter a table name and click OK. DMS creates a table similar to the selected table.
4. The structure of the created table is the same as that of the source table.

   A similar table is created.

14.4.2.5 Generate SQL statement templates

This topic describes how to generate SQL templates in DMS.

Procedure

1. *Log on to an RDS instance through DMS*.
2. In the left-side Table directory tree, right-click the table you want to copy and choose Create SQL Template from the shortcut menu.

3. DMS generates SQL INSERT, UPDATE, SELECT and CREATE TABLE statement templates as a reference when you perform SQL operations.

14.4.2.6 Query table information

This topic describes how to query table information in DMS.

Procedure

1. *Log on to an RDS instance through DMS*.
2. In the left-side Table directory tree, right-click the table you want to query and choose Object Info from the shortcut menu.
3. DMS obtains information about the table object. Click the Basic Info tab to view basic information of the table.
4. Click the Create Statement tab to view the table creation statements.

14.4.2.7 Clear data

This topic describes how to clear table data in DMS.

Procedure

1. *Log on to an RDS instance through DMS*.
2. In the left-side Table directory tree, right-click the table that you want to clear data from and choose Clear Table from the shortcut menu.

   Notice:
   Clearing table data is a high-risk operation and may affect your data usage. DMS prompts you to confirm whether to clear table data.
3. Click Yes if you want to clear table data. DMS then clears data of the selected table.
4. Open the table to check whether its data is cleared.

14.4.2.8 Perform operations on tables in batches

This topic describes how to perform operations on tables in batches in DMS.

Procedure

1. *Log on to an RDS instance through DMS*.

2. Delete tables in batches.
   a) In the left-side Table directory tree, right-click a table and choose Batch Operate Tables > Batch Delete Tables.

      The Batch Delete Tables window appears.
   b) Select the tables to be deleted.
   c) Click OK.

      DMS prompts you to confirm whether you want to delete the selected tables in batches.
   d) Click Yes.

      DMS deletes the selected tables in batches.
3. Perform operations on tables in batches.

   You can clear data, delete or maintain tables, and modify table name prefixes in batches.
   a) In the left-side Table directory tree, right-click a table and choose Operate Tables > More Batch Operations from the shortcut menu.

      The More Batch Operations window appears.
   b) Select the tables to be operated and click Clear Data, Delete, Table Maintenance, or Table Name Prefix.
   c) Click OK.

      DMS prompts you to confirm whether you want to perform the batch operation.
   d) Click Yes.

      DMS performs the batch operation.

14.4.2.9 Maintain a table

This topic describes how to maintain and optimize a table in DMS.

Procedure

1. *Log on to an RDS instance through DMS*.
2. In the left-side Table directory tree, right-click the table you want to maintain and choose Maintain Table > Optimize Table.

3. Click Yes.

   Click Yes if you want to optimize the table. Then DMS starts optimization.

   Optimization allows you to reuse the table space in the database and organize file fragments.

   > **Note:**
   > You can check, restore, and analyze tables in a way similar to optimizing tables.

14.4.3 Manage indexes

This topic describes how to add, modify, or delete indexes in DMS.

Procedure

1. *Log on to an RDS instance through DMS*.
2. In the left-side Table directory tree, expand the table you want to modify and choose Index > Add Index.

   The Add Index page appears.
3. Set index parameters.
   - 1: Enter an index name and select an index type.
   - 2: Click + or — to add or delete a field to or from the index.
   - 3: Edit the fields of the index. You can enter or select values from the drop-down list. You can set a prefix length for a variable-length field (such as varchar) to save space occupied by the index.
4. Click Save.

   DMS generates SQL statements used to add the index. Confirm the change.
5. Click Run.

6. **After the index is added, check the indexes of the table to verify that the new index takes effect.**

   You can modify or delete the new index as needed.

   - **In the left-side Table directory tree, right-click an index and choose Modify Index from the shortcut menu. The Modify Index window appears.**

     The method of modifying an index is similar to that of adding one, except that the SQL statements delete the old index before adding a new one.
   - **In the left-side Table directory tree, right-click an index and choose Delete Index from the shortcut menu. The Delete Index window appears. Click OK to delete the index.**

14.4.4 Manage foreign keys

This topic describes how to add foreign keys in DMS.

Procedure

1. *Log on to an RDS instance through DMS*.
2. **In the left-side Table directory tree, right-click the table to be modified and choose Edit Table Structure from the shortcut menu.**
3. **On the Edit Table page that appears, click the Foreign Key tab to edit foreign keys.**
4. **Enter the foreign key information, and set the fields of foreign keys and referenced tables.**
5. **Click Save.**

14.4.5 Create partitions

This topic describes how to create partitions in DMS.

Procedure

1. *Log on to an RDS instance through DMS*.
2. **In the left-side Table directory tree, right-click a table and choose New Table from the shortcut menu.**

   The Create: Table page is displayed.
3. **Enter the basic table information, and set the table fields and partitions.**

4. Click Save to save the created table structure.

   A window is displayed for you to confirm the SQL statements used to create the table.

5. Click OK. DMS creates the partition table based on the partition fields and partitioning logic that you have configured.
6. After executing the SQL statements, check whether the partition table is created.

14.4.6 Create a stored procedure

This topic describes how to create and manage stored procedures in DMS.

Context

A MySQL database is used as an example.

Stored procedures, functions, triggers, and events are considered programmable objects in DMS.

Procedure

1. *Log on to an RDS instance through DMS*.
2. Click the left-side Programmable Object directory tree, and choose Stored Procedure > Create (Stored Procedure).

   The Create Stored Procedure tab is displayed.

3. Enter a name and a description for the stored procedure.
4. Click OK.
5. DMS provides a template for creating stored procedures. You only need to edit the stored procedure part.
6. Click Save to save the stored procedure to the database.

   If a syntax error is found, DMS returns the cause of the error.

7. Click Run to run the stored procedure.

   DMS displays a page for you to set the input parameters for the stored procedure.

   Set the input parameters. In this example, set `cnt` to `80` to search for records that meet the `Value=80` condition.

8. **Click Execute to execute the stored procedure.**

   DMS displays output parameters or intermediate result set of the stored procedure, if any.

   - **The Message tab displays messages about the execution, such as output variables and intermediate result sets.**
   - **The Intermediate Result Set 1 tab displays the result set generated during the execution. If multiple intermediate result sets are available, DMS will generate multiple tabs, such as Intermediate Result Set 1, Intermediate Result Set 2, and Intermediate Result Set 3.**

9. **Click the Intermediate Result Set 1 tab.**

   DMS displays records with the value of 80.

10. **You can set the options when creating the stored procedure. Click Option Settings to set options for the stored procedure.**
11. **After a stored procedure is created, it is added to the Programmable Object directory tree.**

    You can perform other operations related to the stored procedure through the following menu options:

    - **Create**
    - **Edit**
    - **Delete**
    - **Execute**

12. **You can run the stored procedure in the SQL window.**

    - **1: Run the call stored_procedure_name command to call a stored procedure.**
    - **2: The SQL window shows the result set of the stored procedure, if any.**

14.4.7 Create a function

This topic describes how to create a function in DMS.

Context

Functions, stored procedures, triggers, and events are considered programmable objects in DMS.

Procedure

1. *Log on to an RDS instance through DMS*.

2. In the left-side Programmable Object directory tree, choose Function > Create (Function).

   The Create Function page is displayed.
3. Set basic information of the new function.
4. Click OK.

   The Edit Function page appears. DMS generates a function creation template.
5. Enter information in the function part.
6. Click Save. DMS then checks whether the function is correctly defined. If not, DMS returns an error message.

   DMS runs the correct function definition in your database, and returns a message, indicating that the function is saved.
7. Click Execute to execute the function.
8. Enter a parameter such as `wednesday` and click Execute to execute the function.
9. Click Option Settings to set different options for the function.

   You can also run the function in the SQL window.

14.4.8 Create a view

This topic describes how to create and manage custom views in DMS.

Procedure

Create a view

1. *Log on to an RDS instance through DMS*.
2. Click the View directory tree on the left side to check the views of the current database.
3. Right-click the blank space and choose New View from the shortcut menu.

   The Create: View page is displayed.
4. Set basic information of the view.

   The following example shows how to filter records in the dmstest table whose values are even numbers, and output the id and name fields.
5. Click Save Changes. DMS generates SQL statements used to create the view based on your settings.

6. Click OK after you confirm the SQL statements. DMS saves the defined view to your database.
7. The saved view is added to the View directory tree on the left side. You can click the view to display its definition.

Check the view

8. Right-click View and choose Check View from the shortcut menu to query data through the newly created view.
9. You can perform view-related operations in DMS.

   The menu options include:

   - View Data
   - Create View
   - Edit View
   - Delete View
   - Refresh Views

14.4.9 Create a trigger

This topic describes how to create and manage a trigger in DMS.

Context

Triggers, functions, stored procedures, and events are considered programmable objects in DMS.

Procedure

1. *Log on to an RDS instance through DMS*.

2. Click the Programmable Object directory tree on the left side, and choose Trigger > Create (Trigger).

   The Create: Trigger tab is displayed.

   - 1: Trigger table.
   - 2: Trigger settings.
     - Enter a name for the trigger.
     - Select dmstest from the drop-down list as the trigger table.
     - Select AFTER from the drop-down list as the trigger time.
     - Select INSERT from the drop-down list as the trigger event.
   - 3: Set a trigger statement.
     - Set the operations to be performed when the trigger event occurs.
     - When data is inserted into the dmstest table, the trigger in this example inserts data into the copy_test table and records the insertion time in copy_test.time.
3. Click Save after you finish the trigger settings. DMS then generates the SQL statement to be executed by the trigger based on your settings. Confirm the SQL statement.
4. Click OK. DMS then saves the trigger to your database. DMS returns a message, indicating that the trigger has been saved. In the left-side navigation pane, choose Programmable Object > Trigger to view the trigger you created.
5. You can insert data into the `dmstest` table to check whether the data is recorded in the `copy_test` table.

   1: Insert data into the dmstest table and query the copy_test table for the inserted data.

   2: The SQL window displays messages about the execution of the SQL statements . The messages indicate that a row is inserted into the dmstest table and that this row is also added to the copy_test table.
6. Check the result set in the SQL window to verify whether the insert operation is correctly performed by the trigger.

7. In the left-side navigation pane, choose Programmable Object > Trigger to perform trigger-related operations through the following menu options:
   - Create (Trigger)
   - Edit (Trigger)
   - Delete (Trigger)

14.4.10 Create an event

This topic describes how to create and manage events in DMS.

Prerequisites

After you log on to a database, make sure that event support has been enabled for the database.

- Execute the `SELECT @@event_scheduler;` statement to check whether the database supports events. If ON is returned, event support is enabled.
- If `OFF` is returned, event support is disabled. You need to enable event support by modifying the configuration file or executing the `SET GLOBAL event_scheduler = ON;` statement.

Context

Events, triggers, functions, and stored procedures are considered programmable objects in DMS.

Procedure

1. *Log on to an RDS instance through DMS*.
2. Click the Programmable Object directory tree on the left side, and choose Event > Create (Event).

   The Create Event page is displayed.

   1: In the event setup area, set the event name, cycle, start time, end time, status, and comment, and choose whether to enable cyclic execution.

   2: In the event execution Statement area, set the operations to be performed when a scheduled event is triggered.
3. Set an event trigger rule and the SQL statements for event execution.
4. Click Save. DMS generates the SQL statements used to create the event.

5. **After you confirm that the SQL statements are correct, click OK. DMS then executes the edited event in your database.**
   - **If the event is created, DMS returns a message, indicating that the event is saved.**
   - **In the left-side navigation pane, choose Programmable Object > Event to view the event you created.**
6. **Check whether the event is properly executed in the SQL window.**

   In this example, the event executes SQL statements to insert a piece of data into the `copy_test`table every minute. Check the copy_test table to see whether the data is inserted as programmed.
7. **You can perform various event-related operations in DMS.**

   The menu options include:
   - **Create (Event)**
   - **Edit (Event)**
   - **Delete (Event)**

14.5 Data processing

14.5.1 Import data

This topic describes how to use DMS to import data.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an RDS instance through DMS*.
2. **In the top navigation bar, choose Data Processing > Import.**

   The Import tab is displayed.

   The Import tab contains the import toolbar and import history.

   If you have imported data, you can view previous operations in History.

3. Click New Task.

   The Import File dialog box is displayed.

   - Select the type of the file to be imported. Only SQL and CSV files are supported currently.
   - If the data file uses a character set, you can manually specify the character set . DMS automatically detects character sets of files.
   - DMS terminates the import task if an error occurs while executing an SQL statement. You can select `Ignore Errors` to proceed. However, this operation may affect subsequent operations.
   - You can enter a brief description of the import task for later review.

4. Click Start to start the import task.

   If the imported data has any error, DMS terminates the import process and return an error message. You can modify the data file to correct the error and import it again.

   If the imported data and SQL statements are correct, DMS displays the import progress, volume of the imported data, and time elapsed.

   After the import is completed, you can view the import task in History.

   Click Task Number to view the execution details of the task.

14.5.2 Export data

14.5.2.1 Export a database

This topic describes how to use DMS to export a database.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an RDS instance through DMS*.
2. In the top navigation bar, choose Data Processing > Export to go to the Export page.
3. On the Export page, choose New Task > Export Database.

4. On the Export SQL Result Sets tab, select a database, file type (SQL or CSV), and content to be exported (structure and data, only data, or only structure). Select tables on the right side and additional content in the Additional Content area.
5. Click OK to run the export task.

   DMS refreshes the export progress every two seconds.

   You can close the export window and review the export details, and download the exported data in the Export History List.

   After the export is complete, DMS automatically downloads the exported file to your local computer. You can also click Download File to download the exported file.

   You can view previously submitted export tasks in the Export History List. Click a task name to view the task details and download the exported data.

14.5.2.2 Export an SQL result set

This topic describes how to use DMS to export an SQL result set.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an RDS instance through DMS*.
2. In the top navigation bar, choose Data Processing > Export to go to the Export page.
3. On the Export page, choose Add Task > SQL Result Set Export.
4. On the Export SQL Result Sets tab, complete the settings as needed.

   Select a file type (CSV or SQL_Insert) and a database, set the maximum number of rows of the result set, and enter the SQL statements.

5. Click OK. DMS then runs the SQL result set export task in the background.

   After the export task is completed, DMS automatically downloads the exported files to your local computer. You can also click Download File to download the exported files.

   DMS also summarizes the export results and automatically downloads the exported SQL result set files.

   You can view the SQL result set export tasks that you submitted in the Export History List and download SQL result set files.

14.6 Performance

14.6.1 Lock wait

14.6.1.1 View lock-waits

When an RDS for MySQL session is waiting for an exclusive InnoDB row lock held by another session, InnoDB lock wait will occur. This topic describes how to view lock wait in DMS.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an RDS instance through DMS*.
2. In the top navigation bar, choose Performance > InnoDB Lock Wait.

   If transactions of the current instance are waiting for locks, the lock hold and lock wait are displayed.
3. Move the pointer over the `Lock` or `Lock-Wait` icon to view the locks or lock-waits, and related session IDs.
4. Click  to reload the data.

14.6.1.2 Release lock wait

This topic describes how to release lock wait in DMS.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an RDS instance through DMS*.
2. **In the top navigation bar, choose Performance > InnoDB Lock-Wait.**

   If transactions of the current instance are waiting for locks, the Lock Hold and Lock Wait icons are displayed.
3. **Move the pointer over the `Lock` or `Lock-Wait` icon to view the locks or lock-waits, and related session IDs.**
4. **Click the `Lock` or `Lock-Wait` icon.**

   The Delete Session message appears.
5. **Click Yes to terminate the current session.**

14.7 Extended tools

14.7.1 Table data volume statistics

This topic describes how to view table data volume statistics in DMS.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an RDS instance through DMS*.
2. **In the top navigation bar, choose Tools > Table data amount.**
3. **The page shows the information about all user tables of the current instance, including the database, table name, storage engine, number of rows, row size (in bytes), data, index, creation time, and character set sorting rules.**

   You can filter the statistics on table data volumes based on a range of criteria such as the database name, table name, total table size (in MB), number of table rows, global sorting, and storage engine. You can also perform the paging, refresh, and reset operations.

14.7.2 ER diagrams

This topic describes how to view entity-relationship (ER) diagrams in DMS.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an RDS instance through DMS*.
2. In the top navigation bar, choose Tools > ER Diagram.

   The E-R diagram shows the relationship between the tables of the current database and provides the methods for representing table names, column names , indexes, and relationships.
3. You can perform the following operations:
   - Select another database from the DB:mysql drop-down list.
   - Click the Sorting: Sorting Options drop-down list to sort tables in ascending or descending order of table name, field count, or relationship count.
   - Click Refresh to refresh the current ER diagram.
   - Click View SQL Scripts to list the SQL statements used to create all tables of the current database.
   - Click Download SQL Scripts to download the SQL statements used to create all tables of the current database.
   - Click Download XML Files to download the table-creating SQL statements of the current database in XML format.
   - Double-click the name of a table column to view the column definition.
   - Double-click a table name to edit the table on a new page.

14.8 DMS for Redis

14.8.1 Function overview

Figure 14-15: Homepage **shows the areas on the DMS for Redis homepage.**

Figure 14-15: Homepage

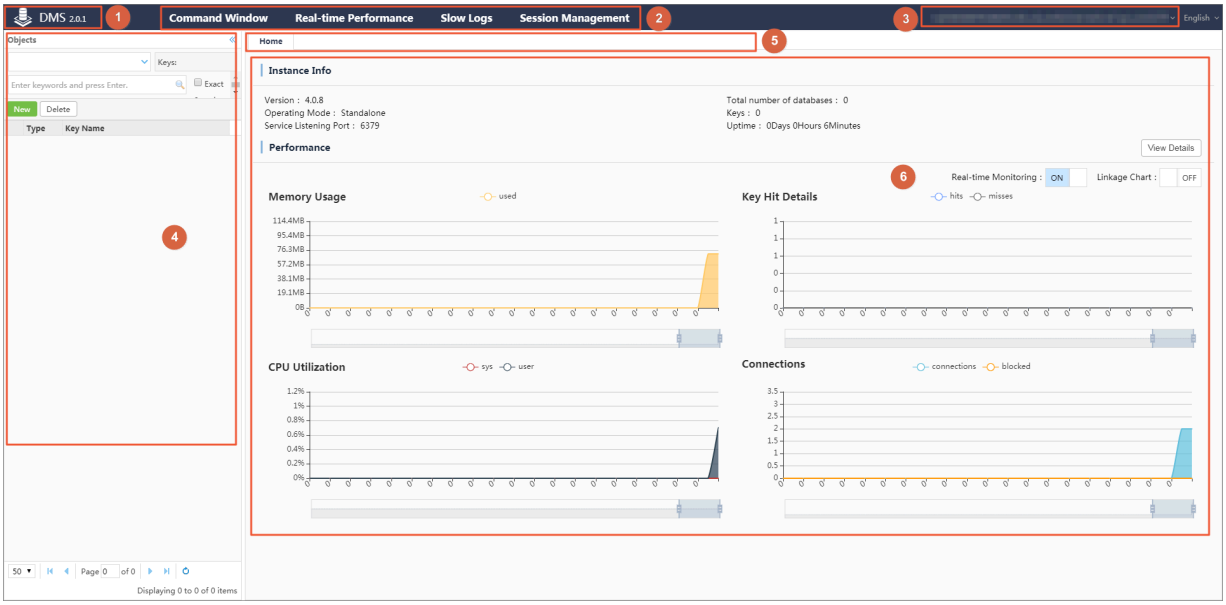


Table 14-7: DMS for Redis homepage description **describes the areas of the homepage shown in the figure.**

Table 14-7: DMS for Redis homepage description

No.	Area	Description
1	Version	You can hover over this area to view the upgrade records of the current version and go to the upgrade history page.
2	Top navigation bar	This area provides multiple functions, including the command window and real-time performance.

No.	Area	Description
3	Instance display area	This area displays the connection string of the current instance. You can hover over the connection string to show the log out menu.
4	Object list	The object list allows you to perform multiple operations. For example, you can select databases, search for keys by keyword, and view search results.
5	Tab	You can click a tab to view the corresponding tab page.
6	Tab page	A tab page displays the basic information and available actions for the corresponding feature.

14.8.2 Data management

14.8.2.1 Create a key

Procedure

1. On the DMS logon page, enter the database logon information and click Log On.
2. After logging on to the instance, select the database for which you want to create a key from the drop-down list.
3. Click New.
4. Set `Key Name` and select a data type of the value from the `Type` drop-down list.

   A value editing tab varies depending on the data types of the value.

5. Click OK. The value editing tab page is displayed.
   - Value type 1: String
     a. Enter a value in the Value area.
     b. Click Commit. A dialog box is displayed, showing the command that will be executed to add the value.
     c. Click OK to create the key.
   - Value type 2: List
     a. Edit the values in the Value list. To add multiple values, click Add to the Head to add values to the head of the list or click Add to the Tail to add values to the tail of the list.
     b. Click Commit. A dialog box is displayed, showing the command that will be executed to add the value.
     c. Click OK to create the key.
   - Value type 3: Hash
     a. Edit the values in the Value list. To add multiple data entries, click New.

        A valid data entry must contain a key and a value. The key must be unique while the same value can be specified for different keys.
     b. Click Commit. A dialog box is displayed, showing the command that will be executed to add the value.
     c. Click OK to create the key.
   - Value type 4: Set
     a. Edit the values in the Value list. To add multiple data entries, click New. The values must be unique.
     b. Click Commit. A dialog box is displayed, showing the command that will be executed to add the value.
     c. Click OK to create the key.
   - Value type 5: ZSet (sorted set)
     a. Edit the values in the Value list. To add multiple data entries, click New.

        A valid data entry must contain a value and a score. The value must be unique while the same score can be specified for different values. A valid score must be an integer or decimal number.

b. **Click Commit. A dialog box is displayed, showing the command that will be executed to add the value.**

c. **Click OK to create the key.**

14.8.2.2 Edit a key

Procedure

1. **On the DMS logon page, enter the database logon information and click Log On.**
2. **After logging on to the instance, select the database that contains the key to be edited from the drop-down list.**
3. **Enter the key name or part of the key name in the search box, and press Enter or click the search icon.**

4. In the search results, double-click the key to be edited. The value editing tab page is displayed.

   A value editing tab varies depending on the data type of the value.

   - Example 1: String

     a. Enter a value in the Value list.

     b. Click Commit. A dialog box is displayed, showing the command that will be executed to edit the value.

     c. Click OK to submit the changes.

   - Example 2: List

     a. Add, edit, or delete values in the Value list.

     - To add multiple values, click Add to the Head to add values to the head of the list or click Add to the Tail to add values to the tail of the list.
     - To edit an existing value, double-click and edit the value.
     - To delete an existing value, select the row containing the value and click Delete.
     - Information about the number of data entries per page and the number of pages is displayed at the lower part of each tab page. You can click the buttons to go to different pages, and locate values as needed.

     b. Click Commit. A dialog box is displayed, showing the command that will be executed to edit the value.

     c. Click OK to submit the changes.

   - Example 3: Hash

     a. Edit the values in the Value list. To add multiple data entries, click New. To edit an existing data entry, double-click and edit the value. To delete an existing data entry, select the row containing the value and click Delete.

      Note:

A valid data entry must contain a key and a value. The key must be unique while the same value can be specified for different keys. You can enter part of a key in the search box to search for the key as required.

b. Click Commit. A dialog box is displayed, showing the command that will be executed to edit the value.

c. Click OK to submit the changes.

- Example 4: Set

a. Edit the values in the Value list. To add multiple data entries, click New. To edit an existing value, double-click and edit the value. To delete an existing value, select the row containing the value and click Delete.

Note:
The values must be unique. You can enter the value or part of the value in the search box to search for the value.

b. Click Commit. A dialog box is displayed, showing the command that will be executed to edit the value.

c. Click OK to submit the changes.

- Example 5: ZSet (sorted set)

a. Edit the values in the Value list. To add multiple data entries, click New. To edit an existing data entry, double-click and edit the value. To delete an existing data entry, select the row containing the value and click Delete.

Note:
A valid data entry must contain a value and a score. The value must be unique while the same score can be specified for different values. A valid score must be an integer or decimal number. You can search for a specified value. Click Switch search to switch between search by keyword or search by score range.

b. Click Commit. A dialog box is displayed, showing the command that will be executed to edit the value.

c. Click OK to submit the changes.

14.8.2.3 Set key timeout

Procedure

1. On the DMS logon page, enter the database logon information and click Log On.
2. After logging on to the instance, select the database that contains the key to be edited from the drop-down list.
3. Enter the key name or part of the key name in the search box, and press Enter or click the search icon.
4. In the search results, locate the key to be edited. Right-click the key and choose Set Timeout from the shortcut menu.
5. In the displayed dialog box, enter a value for Set Timeout (unit: second).

   Note:

   - We recommend that you do not manually enter –1 in Set Timeout. A key with a value less than 0 immediately expires and cannot be searched for.
   - If Set Timeout is -1 by default, the timeout of the key has not been set, and the key will not expire. The displayed timeout value is consistent with the output of the TTL command in a Redis database.
6. In the displayed message, click Yes to complete the timeout setting.

   Note:
   An expired key cannot be searched for.

14.8.2.4 Delete a key

Procedure

1. On the DMS logon page, enter the database logon information and click Log On.
2. After logging on to the instance, select the database that contains the key to be deleted from the drop-down list.
3. Enter the key name or part of the key name in the search box, and press Enter or click the search icon.
4. In the search results, locate the key to be deleted, right-click the key, and choose Delete from the shortcut menu.
5. In the displayed message, click Yes to delete the key.

14.8.2.5 Rename a key

Procedure

1. On the DMS logon page, enter the database logon information and click Log On.

2. **After logging on to the instance, select the database that contains the target key from the drop-down list.**
3. **Enter the key name or part of the key name in the search box, and press Enter or click the search icon.**
4. **In the search results, right-click the key to be renamed and choose Rename from the shortcut menu.**
5. **Enter a new key name.**
6. **Click Yes to rename the key.**

14.8.3 Performance monitoring

14.8.3.1 View the homepage

Procedure

1. **On the DMS logon page, enter the database logon information and click Log On. The homepage is displayed.**
2. **View the homepage.**
   - **On the homepage, basic instance information is displayed in the upper part, and performance metrics are displayed in the lower part.**
   - **Real-time monitoring data collection starts when you open the homepage. The data on the page is refreshed every eight seconds. The refresh interval cannot be changed. You can turn the Real-Time Monitoring switch on or off to enable or disable monitoring data refresh.**
   - **You can hover over a chart to view the data collected at a specific point in time.**
   - **You can turn the Linkage Chart switch on or off to enable or disable the linkage among charts. If you turn on the Linkage Chart switch and then hover**

over one of the charts, all charts display the data collected at the same specific point in time.

- **In most cases, there is a slider indicating time period below a chart. You can adjust both ends of the slider to view the data collected within the adjusted time period.**
- **Some charts contain multiple metrics. The metric legends are displayed above each chart. Each metric legend corresponds to a line in a chart in the same color. Click a metric legend to display or hide the metric in the chart.**
- **Click Show Details to go to the real-time performance page. You can also click Real-time Performance in the top navigation bar to go to the page.**

14.8.3.2 View real-time performance

Procedure

1. **On the DMS logon page, enter the database logon information and click Log On.**

2. **Click Real-Time Performance in the top navigation bar. The real-time performance page is displayed.**

   - **On the real-time performance page, metric data is displayed in rectangular boxes in the upper real-time data area, and metric data trends in different charts are displayed in the lower chart area.**
   - **Real-time monitoring data collection starts when you open the homepage. The data on the page is refreshed every eight seconds. The refresh interval cannot be changed.**
   - **You can turn the Real-Time Monitoring switch on or off to enable or disable monitoring data refresh.**

   In the real-time data area:

   - **Each metric (displayed in a rectangular box) is correlated to a chart in the lower part of the page. Click a rectangular box to display or hide the correlated chart.**
   - **If the color of a rectangular box is blue, the correlated chart is displayed. Otherwise, the chart is hidden.**

   In the chart area:

   - **You can hover over a chart to view the data collected at a specific point in time.**
   - **In most cases, there is a slider indicating time period below a chart. You can adjust both ends of the slider to view the data collected within the adjusted time period.**
   - **Some charts contain multiple metrics. The metric legends are displayed above each chart. Each metric legend corresponds to a line in a chart in the same color. Click a metric legend to display or hide the metric in the chart.**

14.9 DMS for MongoDB

14.9.1 Function overview

Figure 14-16: Homepage **shows the areas on the DMS for MongoDB homepage.**

Figure 14-16: Homepage

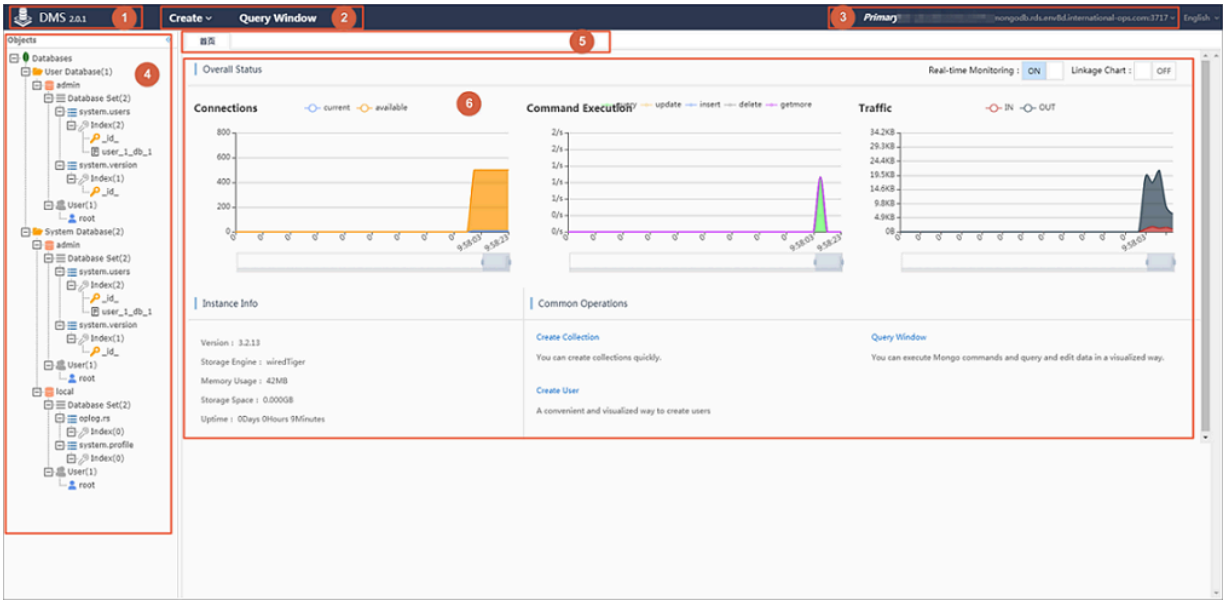


Table 14-8: DMS for MongoDB homepage description **describes the areas of the homepage shown in the figure.**

Table 14-8: DMS for MongoDB homepage description

No.	Area	Description
1	Version	You can hover over this area to view the upgrade records of the current version and go to the upgrade history page.
2	Top navigation bar	This area provides multiple functions such as database creation, collection creation, user creation and the query window.
3	Instance display area	This area displays the connection string of the current instance. You can hover over the connection string to show the log out menu.
4	Object list	The object list displays the structure of database objects, including databases, collections, users, and indexes. You can right-click these objects to manage them.
5	Tab	You can click a tab to view the corresponding tab page.

No.	Area	Description
6	Tab page	A tab page displays the basic information and available actions for the corresponding feature.

14.9.2 Structure management

14.9.2.1 Create a collection

Procedure

1. On the DMS logon page, enter the database logon information and click Log On.
2. From the top navigation bar, choose Create > Database Set.

   The Create Collection dialog box is displayed.
3. Enter the name of the database for which you want to create a collection, and enter a collection name.
4. Click Yes to create the collection.

14.9.2.2 Create a database

Procedure

1. On the DMS logon page, enter the database logon information and click Log On.
2. From the top navigation bar, choose Create > Database.
3. Enter a database name and collection name.

   Note:
   When you create a database, you must enter a collection name in the Create Database dialog box to create a collection for the database. If you do not enter a collection name, a collection named test is created by default.
4. Click Yes to create the database.

14.9.2.3 Create an index

Procedure

1. On the DMS logon page, enter the database logon information and click Log On.
2. Expand the left-side object list and locate the collection for which you want to create an index.
3. Right-click the collection and choose Create Index from the shortcut menu. The Add Index dialog box is displayed.

4. In Add Index, enter an index name, add an index key, and specify a sorting order for the key. You can click New to add multiple keys. Make sure you specify a proper sorting order for the index keys.

   Note:
   Some parameters are displayed on the Advanced Options tab page. Click the Advanced Options tab to configure them.

5. Click Yes to complete the index creation.

14.9.2.4 Edit an index

Procedure

1. On the DMS logon page, enter the database logon information and click Log On.
2. Expand the left-side object list and locate the index to be edited.
3. Right-click the index and choose Edit Index from the shortcut menu.

   To change the index name, you can enter a new name in the Index Name field. To manage index keys, you can add, delete, or modify the keys in the Index area. Make sure you specify a proper sorting order for the index keys.

   Some parameters are displayed on the Advanced Options tab page. Click the Advanced Options tab to configure them.

4. Click Yes to complete the index configurations.

14.9.2.5 Delete a collection

Procedure

1. On the DMS logon page, enter the database logon information and click Log On.
2. Expand the left-side object list and locate the collection to be deleted. Right-click the collection and choose Drop Collection from the shortcut menu.
3. In the displayed Information message, click Yes to delete the collection.

14.9.2.6 Delete a database

Procedure

1. On the DMS logon page, enter the database logon information and click Log On.
2. Expand the left-side object list and locate the database to be deleted. Right-click the database and choose Drop Database from the shortcut menu.
3. In the displayed message, click Yes to delete the database.

Related tasks
Delete a collection

14.9.2.7 Delete an index

Procedure

1. **On the DMS logon page, enter the database logon information and click Log On.**
2. **Expand the left-side object list and locate the index to be deleted.**
3. **Right-click the index and choose Drop Index from the shortcut menu.**
4. **In the displayed message, click Yes to delete the index.**

14.9.3 User management

14.9.3.1 Create a user

Procedure

1. **On the DMS logon page, enter the database logon information and click Log On.**
2. **From the top navigation bar, choose Create > Users.**

   The Create User dialog box is displayed.

   - **If you have logged on to the admin database, you can create a user with the highest privilege.**
   - **You can click the Privileges on Other Databases tab, select a database, and assign permissions on the selected database to the user.**
3. **Click the Privileges on Current Database tab, select the permissions that you want to assign to the user, and click Yes.**

14.9.3.2 Edit a user

Procedure

1. **On the DMS logon page, enter the database logon information and click Log On.**
2. **Expand the left-side object list and locate the user to be edited. Right-click the user and choose Edit User from the shortcut menu.**
3. **On the Privileges on Current Database tab page of the displayed dialog box, select the permissions that you want to assign to the user.**

   If the user belongs to the admin database, you can click the Privileges on Other Databases tab, select a database, and assign permissions on the selected database to the user.

4. **Click Yes.**

14.9.3.3 Delete a user

Procedure

1. **On the DMS logon page, enter the database logon information and click Log On.**
2. **Expand the left-side object list and locate the user to be deleted. Right-click the user and choose Drop User from the shortcut menu.**
3. **In the displayed message, click Yes to delete the user.**

14.9.4 Data management

14.9.4.1 Create a document

Procedure

1. **On the DMS logon page, enter the database logon information and click Log On.**
2. **Expand the left-side object list and locate the collection for which you want to create a document. Right-click the collection and choose View Data from the shortcut menu.**
3. **In the displayed Query Window, a query command is executed automatically to query the documents in the collection. Click Create Document on the Results tab page.**

   A dialog box for creating a document is displayed.

   - **Enter the document content in the displayed dialog box and ensure the content is compliant with the mongo shell standards.**
   - **You can choose whether to enclose an element name in double quotation marks (" "). If the element name contains spaces, you must enclose the name in double quotation marks (" ").**
4. **After you edit the content, click Validate Format to validate the format in the document.**

   If a message indicating the format is valid is displayed, the document has passed the format validation. Otherwise, you need to modify the content based on the error message.
5. **After the format validation succeeds, click Yes in the dialog box. The Confirm Commands dialog box is displayed. Confirm the entered content and click Yes in the Confirm Commands dialog box.**

14.9.4.2 Edit a document

Procedure

1. **On the DMS logon page, enter the database logon information and click Log On.**
2. **After logging on to the instance, expand the left-side object list and locate the collection that contains the document to be edited. Right-click the collection and choose View Data from the shortcut menu.**
3. **In the displayed Query Window, a query command is executed automatically t to query the documents in the collection. You can modify the query command to add query conditions and find the document to be edited.**
4. **You can edit the elements in the document or directly edit the document.**

   - **Edit an element in a document**

     You can edit elements when the changes do not affect the document structure.

     a. **Locate the element to be edited and click the value field. If the element can be edited, it enters the edit mode.**

        The edit mode varies with the data type of the element. For an element of the time type, a calendar is displayed. For an element of the boolean type, a drop-down list that contains boolean values is displayed. For an element of the string type, a text box is displayed.

     b. **After editing the element, click another field to complete editing the current element. In the displayed Confirm Commands dialog box, confirm the element modification and click Yes.**

        Right-click an element and choose Delete Document from the shortcut menu. In the displayed Confirm Commands message, click Yes to delete the element.

   - **Directly edit a document**

     The procedure of directly editing a document is similar to that of creating a document. You can replace the content in a document to edit the document. You can edit a document when you need to make a large number of changes to

the document or make changes to the document structure, such as adding or deleting elements.

a. Select the document to be edited in the List view and locate an element in the document. Right-click the element and choose Edit Document from the shortcut menu.
b. Edit the document content in the displayed dialog box and ensure the content is compliant with the mongo shell standards.
   - You can choose whether to enclose an element name in double quotation marks (" ").
   - If the element name contains spaces, you must enclose the name in double quotation marks (" ").
   - You can delete an element in the Edit dialog box.
   - You can add an element by editing a document or adding the element in the Edit dialog box.

    **Note:**
   When you edit an element, follow the format specified for the data type of the element. If you change the format, the data type of the element may also change. For example, `Value: NumberInt(123)` indicates that the data type of the value is integer. `Value: 123` indicates that the data type of the value is double. If you edit the element and confirm the change, the data type of the value changes from integer to double.
c. After you edit the content, click Validate Format to validate the format in the document. If a message indicating *`the format is valid`* is displayed, the document has passed the format validation. Otherwise, you need to modify the content based on the error message.
d. After the format validation succeeds, click Yes in the Edit dialog box. The Confirm Commands dialog box is displayed. Confirm the changes and click Yes in the Confirm Commands dialog box.

14.9.4.3 Query a document

Procedure

1. On the DMS logon page, enter the database logon information and click Log On.
2. After logging on to the instance, click Query Window in the top navigation bar.

3. Select the database that contains the target document from the Database drop-down list, enter the query command on the command line, and click Execute (F8) to execute the command.

   Note:

   The command output is displayed on the Results tab page in the lower part of the Query Window. If you execute multiple query commands, all command output is displayed on the Results tab page in the execution order of the commands.

   The structure of documents is complex and similar to that of a JSON file. Therefore, DMS provides three document views. You can click buttons on the left side to switch between views as needed.

   - JSON view

     The JSON view is the default view of a document. You can click – or + to the left of an array or object to collapse or expand the structure of the array or object.
   - List view

     The List view displays a document as a tree table. You can click the arrow displayed to the left of an array or object that contains data elements to collapse or expand the structure of the array or object. The List view displays detailed information about each data element, such as data type and value. You can also edit a document in the List view.
   - Text view

     The Text view displays a document in the JSON text format with indents, which allows you to quickly copy the content.

14.9.4.4 Delete a document

Procedure

1. On the DMS logon page, enter the database logon information and click Log On.
2. After logging on to the instance, expand the left-side object list and locate the collection that contains the document to be deleted. Right-click the collection and choose View Data from the shortcut menu.

3. **In the displayed Query Window, a query command is executed automatically to query the documents in the collection. You can modify the query command to add query conditions and find the document to be deleted. Click List on the Results tab page.**
4. **Select a document in the List view, and click Delete Document.**
5. **In the displayed message, click Yes to delete the document.**

14.9.5 View the homepage

Procedure

1. **On the DMS logon page, enter the database logon information and click Log On.**
2. **After logging on to the instance, the homepage is displayed by default.**
   - **On the homepage, performance metrics are displayed in the upper part, and basic instance information is displayed in the lower part.**
   - **Real-time monitoring data collection starts when you open the homepage. The data on the homepage is refreshed every eight seconds. The refresh interval cannot be changed. You can turn the Real-Time Monitoring switch on or off to enable or disable monitoring data refresh.**
   - **You can hover over a chart to view the data collected at a specific point in time.**
3. **You can turn the Linkage Chart switch on or off to enable or disable the linkage among charts. If you turn on the Linkage Chart switch and then hover over one of the charts, all charts display the data collected at the same specific point in time.**
4. **You can adjust both ends of the slider below each chart to view the data collected within the adjusted time period.**

   Some charts contain multiple metrics. The metric legends are displayed above each chart. Each metric legend corresponds to a line in a chart in the same color. Click a metric legend to display or hide the metric in the chart.

15 KVStore for Redis

15.1 What is KVStore for Redis?

KVStore for Redis is an online key-value storage service compatible with open-source Redis protocols. KVStore for Redis supports various types of data, such as strings, lists, sets, sorted sets, and hash tables. The service also supports advanced features, such as transactions, message subscription, and message publishing.

You can easily deploy and manage KVStore for Redis databases in the KVStore for Redis console.

- **You can create an instance to initialize a database environment.**
- **Before using a KVStore for Redis instance, you must add one or more IP addresses or CIDR blocks that you use to connect to databases to the whitelist of the instance.**
- **You can manage instances in the KVStore for Redis console.**
- **To secure data, you can periodically or immediately back up or restore databases in the KVStore for Redis console.**
- **You can log on to a database by using a client and then use SQL statements to perform database operations.**

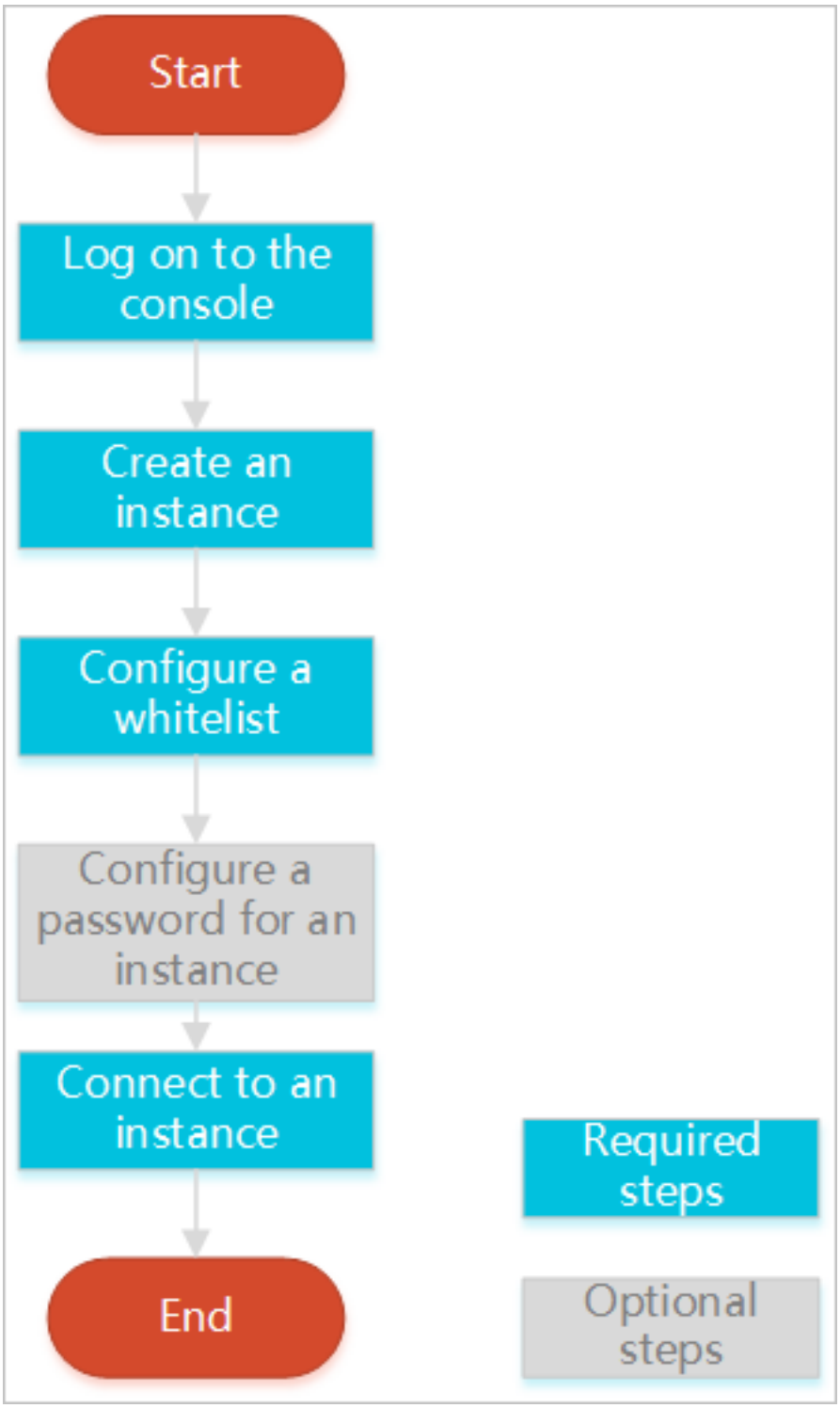
15.2 Quick start

15.2.1 Get started with KVStore for Redis

This topic describes a series of operations from creating a KVStore for Redis instance to logging on to a database. In this way, you can easily understand the procedure of using the KVStore for Redis instance.

For more information about the procedure, see *Figure 15-1: Flowchart for the KVStore for Redis instance*.

Figure 15-1: Flowchart for the KVStore for Redis instance



- **Log on to the** *KVStore for Redis console*

  This topic describes how to log on to the KVStore for Redis console.

- *Create an instance*

  KVStore for Redis supports two types of networks: classic network and Virtual Private Cloud (VPC). You can create KVStore for Redis instances in different networks.

- *Set a whitelist*

  To ensure database security and stability, before using a KVStore for Redis instance, you must add one or more IP addresses or CIDR blocks that you use to connect to databases to the whitelist of the instance.

- *Reset a password*

  If you have not specified a password when creating the instance, set the password of the instance on the Instance Information page.

- *Connect to the instance*

  You can use a client that supports Redis protocols or use the Redis command-line interface (redis-cli) program to connect to the KVStore for Redis instance.

15.2.2 Log on to the KVStore for Redis console

This topic describes how to log on to the KVStore for Redis console.

Prerequisites

- **Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://`IP address or domain name of the Apsara Stack console`/manage.**
- **We recommend that you use the Chrome browser.**

Procedure

1. **Open your browser.**
2. **In the address bar, enter the access address of the Apsara Stack console in the format of http://`IP address or domain name of the Apsara Stack console`/manage, and then press Enter.**

3. Enter the correct username and password.
   - The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click LOGIN to go to the Dashboard page.
5. On the menu bar, choose  > Database > KVStore.
6. Click the KVStore for Redis tab.

15.2.3 Create an instance

This topic describes how to create an instance in the KVStore for Redis console.

Prerequisites

- You have requested an Alibaba Cloud account for logging on to the KVStore for Redis console.
- To use the Virtual Private Cloud (VPC) service, you must create a VPC in the same region as KVStore for Redis.

> **Notice:**
> You must specify the network type when you create the instance, and cannot modify the network type later.

Procedure

1. Log on to the *KVStore for Redis console*.

2. **On the KVStore for Redis tab page, click Create Instance. On the Create Redis Instance page that appears, set parameters as required.**

   Before you select the VPC type, create a VPC. For more information, see the "Create a default VPC and VSwitch" topic of *VPC User Guide* **.**

   Table 15-1: KVStore for Redis parameters

Field	Parameter	Description
Region	**Region**	**Specifies the region where the target KVStore for Redis instance is located.**
	Zone	**Specifies the zone where the target KVStore for Redis instance is located.**
Basic Settings	**Department**	**Specifies the department that the target KVStore for Redis instance belongs to.**
	Project	**Specifies the project that the target KVStore for Redis instance belongs to.** **Notice:** **After you specify the project, only the members that belong to this project can use this instance. For more information, see the "View project members" topic of** *KVStore for Redis User Guide* **.**
Engine Information	**Engine Version**	**Specifies the engine version of the target KVStore for Redis instance.** **You can select Redis 2.8 or Redis 4.0 as the engine version.**

Field	Parameter	Description
Instance Specification	Architecture	Specifies an architecture type for the target KVStore for Redis instance. KVStore for Redis provides cluster and standard architectures. The cluster architecture supports large-capacity or high-performance requirements. Due to Redis single-thread mechanism, we recommend that you use a standard architecture if your business requires QPS performance of 100,000 or less. To require higher performance, select a cluster architecture.
	Node Type	Specifies a node type for the target KVStore for Redis instance. KVStore for Redis supports the dual-copy structure.
	Service Plan	Specifies a standard or premium plan. A premium plan supports advanced editions.
	Instance Specification	Specifies the specifications for the target KVStore for Redis instance. The maximum number of connections and maximum internal network bandwidth vary according to different instance specifications.

Field	Parameter	Description
Network	Network Type	On the Alibaba Cloud platform, a classic network and a VPC have thefollowing differences: · Classic network: cloud services in a classic network are not isolated. Unauthorized access to a cloud service is blocked only by a security group or a whitelist policy of the service. · VPC: a VPC helps you build an isolated network environment in Alibaba Cloud. You can customize the route table, CIDR blocks, and gateway of a VPC. In addition, to smoothly migrate applications to the cloud, you can use a leased line or virtual private network (VPN) to integrate an on-premises data center and cloud resources in a VPC into a virtual data center. Note: Before you select the VPC type, create a VPC. For more information, see the "Create a default VPC and VSwitch" topic of *VPC User Guide* .
Password	Set Password	Specifies the password for connecting to the target instance. Select Now to start setting the password, or select Later to set the password after creating the instance. For more information about how to set the password, see *Reset a password*. A password must follow these rules: · The password must be 8 to 30 characters in length. · The password must contain uppercase and lowercase letters and numbers at the same time, and does not support special characters.
Instance Name	Instance Name	Specifies the name of the target instance. · An instance name must be 2 to 128 characters in length. · The name must start with an uppercase or lowercase letter or a Chinese character and can contain letters, numbers, underscores (_), and hyphens(-).

3. Click Create.

   Afterward, wait until the instance stays in Normal status.

15.2.4 Set a whitelist

To ensure database security and stability, before using a KVStore for Redis instance, you must add one or more IP addresses or CIDR blocks that you use to connect to databases to the whitelist of the instance.

Context

We recommend that you periodically check and adjust your whitelist to improve the access security protection and secure data in KVStore for Redis.

Procedure

1. Log on to the *KVStore for Redis console*.
2. On the KVStore for Redis tab page, click the target instance ID or choose > View Details in the Actions column next to the target instance to go to the Instance Information tab page.
3. On the Instance Information tab page, click Change Whitelist.
4. In the Change Whitelist dialog box that appears, set parameters as required.

   Enter one or more IP addresses or CIDR blocks that you use to connect to the KVStore for Redis instance. To allow connections from all IP addresses or CIDR blocks, set the whitelist to 0.0.0.0/0. To block connections from all IP addresses or CIDR blocks, set the whitelist to 127.0.0.1. We recommend that you delete the default IP address 127.0.0.1. Otherwise, the IP addresses or CIDR blocks that you have added do not take effect.

   Notice:
   When you enter multiple IP addresses or CIDR blocks, separate them with commas (,) and leave no space before or after each comma, such as 192.168.0.1,172.16.213.9. KVStore for Redis supports a whitelist that contains a maximum of 1,000 IP addresses or CIDR blocks.

5. Click OK.

15.2.5 Connect to an instance

15.2.5.1 Use a Redis client

15.2.5.1.1 Overview

You can connect to KVStore for Redis by using clients that support Redis protocols.

The KVStore for Redis service is fully compatible with the native Redis database service. You can connect to both database services in similar ways. Any clients compatible with Redis protocols can connect to KVStore for Redis. You can select a Redis client based on your application features.

Note:

KVStore for Redis only supports connections over an internal network. Therefore, only the ECS instances that run on the same node as KVStore for Redis and that are installed with Redis clients can connect to KVStore for Redis for data operations.

For more information about Redis clients, see *http://redis.io/clients*.

15.2.5.1.2 Jedis client

This topic describes how to connect to a KVStore for Redis instance by using a Jedis client.

Download a Jedis client

For more information, click *Jedis*.

Example of single Jedis connection

```
import redis.clients.jedis.Jedis;
public class jedistest {
public static void main(String[] args) {
    try {
        String host = "xx.kvstore.aliyuncs.com";//You can view the
connection address of the target instance in the console.
        int port = 6379;
        Jedis jedis = new Jedis(host, port);
        //Authentication information
        jedis.auth("password");//password
        String key = "redis";
        String value = "aliyun-redis";
        //Select a database. Default value: 0.
        jedis.select(1);
        //Configure a key.
        jedis.set(key, value);
        System.out.println("Set Key " + key + " Value: " + value);
        //Obtain the configured key value.
        String getvalue = jedis.get(key);
        System.out.println("Get Key " + key + " ReturnValue: " +
getvalue);
        jedis.quit();
        jedis.close();
    } catch (Exception e) {
        e.printStackTrace();
    }
}
```

```
}
```

Example of JedisPool

Configuration file

You can configure the pom configuration file based on the specified client version. Follow these configurations:

```
<dependency>
<groupId>redis.clients</groupId>
<artifactId>jedis</artifactId>
<version>2.7.2</version>
<type>jar</type>
<scope>compile</scope>
</dependency>
```

References to be added

```
import org.apache.commons.pool2.PooledObject;
import org.apache.commons.pool2.PooledObjectFactory;
import org.apache.commons.pool2.impl.DefaultPooledObject;
import org.apache.commons.pool2.impl.GenericObjectPoolConfig;
import redis.clients.jedis.HostAndPort;
import redis.clients.jedis.Jedis;
import redis.clients.jedis.JedisPool;
import redis.clients.jedis.JedisPoolConfig;
```

Example of Jedis-2.7.2

```
JedisPoolConfig config = new JedisPoolConfig();
//Maximum number of idle connections. You can customize this
parameter. Make sure that the specified maximum number of idle
 connections does not exceed the maximum number of connections
that the KVStore for Redis instance supports.
config.setMaxIdle(200);
//Maximum number of connections. You can customize this
parameter. Make sure that the specified maximum number of
connections does not exceed the maximum number of connections
that the KVStore for Redis instance supports.
config.setMaxTotal(300);
config.setTestOnBorrow(false);
config.setTestOnReturn(false);
String host = "*.aliyuncs.com";
String password = "Password";
JedisPool pool = new JedisPool(config, host, 6379, 3000,
password);
Jedis jedis = null;
try {
    jedis = pool.getResource();
    /// ... do stuff here ... for example
    jedis.set("foo", "bar");
    String foobar = jedis.get("foo");
    jedis.zadd("sose", 0, "car");
    jedis.zadd("sose", 0, "bike");
    Set<String> sose = jedis.zrange("sose", 0, -1);
} finally {
```

```
    if (jedis ! = null) {
        jedis.close();
    }
}
/// ... when closing your application:
pool.destroy();
```

Examples of Jedis-2.6 and Jedis-2.5

```
JedisPoolConfig config = new JedisPoolConfig();
//Maximum number of idle connections. You can customize this
parameter. Make sure that the specified maximum number of idle
 connections does not exceed the maximum number of connections
that the KVStore for Redis instance supports.
config.setMaxIdle(200);
//Maximum number of connections. You can customize this
parameter. Make sure that the specified maximum number of
connections does not exceed the maximum number of connections
that the KVStore for Redis instance supports.
config.setMaxTotal(300);
config.setTestOnBorrow(false);
config.setTestOnReturn(false);
String host = "*.aliyuncs.com";
String password = "Password";
JedisPool pool = new JedisPool(config, host, 6379, 3000,
password);
Jedis jedis = null;
boolean broken = false;
try {
    jedis = pool.getResource();
    /// ... do stuff here ... for example
    jedis.set("foo", "bar");
    String foobar = jedis.get("foo");
    jedis.zadd("sose", 0, "car");
    jedis.zadd("sose", 0, "bike");
    Set<String> sose = jedis.zrange("sose", 0, -1);
} catch(Exception e) {
    broken = true;
} finally {
    if (broken) {
        pool.returnBrokenResource(jedis);
    } else if (jedis ! = null) {
        pool.returnResource(jedis);
    }
}
```

15.2.5.1.3 phpredis client

This topic describes how to connect to a KVStore for Redis instance by using a phpredis client.

Download a phpredis client

For more information, click *phpredis*.

Sample code

```
<? php
  /* Replace the following parameter values with the host name and the
 port number of the target instance. */
  $host = "localhost";
  $port = 6379;
  /* Replace the following parameter value with the ID and password of
 the target instance. */
  $user = "test_username";
  $pwd = "test_password";
  $redis = new Redis();
  if ($redis->connect($host, $port) == false) {
    die($redis->getLastError());
  }
  if ($redis->auth($pwd) == false) {
    die($redis->getLastError());
  }
  /* You can perform database operations after authentication. For
more information, see https://github.com/phpredis/phpredis. */
  if ($redis->set("foo", "bar") == false) {
    die($redis->getLastError());
  }
  $value = $redis->get("foo");
  echo $value;
? >
```

15.2.5.1.4 redis-py client

This topic describes how to connect to a KVStore for Redis instance by using a redis-py client.

Download a redis-py client

For more information, click *redis-py*.

Sample code

```
#! /usr/bin/env python
#-*- coding: utf-8 -*-
import redis
#Replace the following parameter values with the host name and the
port number of the target instance.
host = 'localhost'
port = 6379
#Replace the following parameter value with the password of the target
 instance.
pwd = 'test_password'
r = redis.StrictRedis(host=host, port=port, password=pwd)
#You can perform database operations after you establish a connection
. For more information, see https://github.com/andymccurdy/redis-py.
r.set('foo', 'bar');
```

```
print r.get('foo')
```

15.2.5.1.5 C or C++ client

This topic describes how to connect to a KVStore for Redis instance by using the C or C++ program.

The following example describes how to use a KVStore for Redis instance based on the C or C++ program.

Download, compile, and install a C client

```
git clone https://github.com/redis/hiredis.git
cd hiredis
make
sudo make install
```

Write test code

```
 #include <stdio.h>
 #include <stdlib.h>
 #include <string.h>
 #include <hiredis.h>
 int main(int argc, char **argv) {
     unsigned int j;
     redisContext *c;
     redisReply *reply;
     if (argc < 4) {
         printf("Usage: example xxx.kvstore.aliyuncs.com 6379
instance_id password\n");
         exit(0);
     }
     const char *hostname = argv[1];
     const int port = atoi(argv[2]);
     const char *instance_id = argv[3];
     const char *password = argv[4];
     struct timeval timeout = { 1, 500000 }; // 1.5 seconds
     c = redisConnectWithTimeout(hostname, port, timeout);
     if (c == NULL || c->err) {
         if (c) {
             printf("Connection error: %s\n", c->errstr);
             redisFree(c);
         } else {
             printf("Connection error: can't allocate redis context\n
");
         }
         exit(1);
     }
     /* AUTH */
     reply = redisCommand(c, "AUTH %s", password);
     printf("AUTH: %s\n", reply->str);
     freeReplyObject(reply);
     /* PING server */
     reply = redisCommand(c,"PING");
     printf("PING: %s\n", reply->str);
     freeReplyObject(reply);
     /* Set a key */
     reply = redisCommand(c,"SET %s %s", "foo", "hello world");
     printf("SET: %s\n", reply->str);
```

```
    freeReplyObject(reply);
    /* Set a key using binary safe API */
    reply = redisCommand(c,"SET %b %b", "bar", (size_t) 3, "hello", (
size_t) 5);
    printf("SET (binary API): %s\n", reply->str);
    freeReplyObject(reply);
    /* Try a GET and two INCR */
    reply = redisCommand(c,"GET foo");
    printf("GET foo: %s\n", reply->str);
    freeReplyObject(reply);
    reply = redisCommand(c,"INCR counter");
    printf("INCR counter: %lld\n", reply->integer);
    freeReplyObject(reply);
    /* again ... */
    reply = redisCommand(c,"INCR counter");
    printf("INCR counter: %lld\n", reply->integer);
    freeReplyObject(reply);
    /* Create a list of numbers, from 0 to 9 */
    reply = redisCommand(c,"DEL mylist");
    freeReplyObject(reply);
    for (j = 0; j < 10; j++) {
        char buf[64];
        snprintf(buf,64,"%d",j);
        reply = redisCommand(c,"LPUSH mylist element-%s", buf);
        freeReplyObject(reply);
    }
    /* Let's check what we have inside the list */
    reply = redisCommand(c,"LRANGE mylist 0 -1");
    if (reply->type == REDIS_REPLY_ARRAY) {
        for (j = 0; j < reply->elements; j++) {
            printf("%u) %s\n", j, reply->element[j]->str);
        }
    }
    freeReplyObject(reply);
    /* Disconnects and frees the context */
    redisFree(c);
    return 0;
 }
```

Compile the code

```
gcc -o example -g example.c -I /usr/local/include/hiredis -lhiredis
```

Test the code

```
example xxx.kvstore.aliyuncs.com 6379 instance_id password
```

15.2.5.1.6 .NET client

This topic describes how to connect to a KVStore for Redis instance by using the .NET program.

The following example describes how to use a KVStore for Redis instance based on the .NET program.

1. **Download and use a .NET client.**

```
git clone https://github.com/ServiceStack/ServiceStack.Redis
```

2. **Create a *.NET* project.**
3. **Add the reference file stored in the library file directory *ServiceStack.Redis/lib/tests* to the client.**

Sample code

```
using System;
     using System.Collections.Generic;
     using System.Linq;
     using System.Text;
     using System.Threading.Tasks;
     using ServiceStack.Redis;
     namespace ServiceStack.Redis.Tests
     {
         class Program
         {
             public static void RedisClientTest()
             {
                 string host = "127.0.0.1";/*IP address of the host
that you want to connect to*/
                 string password = "password";/*Password*/
                 RedisClient redisClient = new RedisClient(host, 6379
, password);
                 string key = "test-aliyun";
                 string value = "test-aliyun-value";
                 redisClient.Set(key, value);
                 string listKey = "test-aliyun-list";
                 System.Console.WriteLine("set key " + key + " value "
 + value);
                 string getValue = System.Text.Encoding.Default.
GetString(redisClient.Get(key));
                 System.Console.WriteLine("get key " + getValue);
                 System.Console.Read();
             }
             public static void RedisPoolClientTest()
             {
                 string[] testReadWriteHosts = new[] {
                 "redis://password@127.0.0.1:6379"/*redis://Password@
IP address that you want to connect to:Port*/
                 };
                 RedisConfig.VerifyMasterConnections = false;//You
must set the parameter.
                 PooledRedisClientManager redisPoolManager = new
PooledRedisClientManager(10/*Number of connections in the pool*/, 10/*
Connection pool timeout value*/, testReadWriteHosts);
                 for (int i = 0; i < 100; i++)
                 {
                     IRedisClient redisClient = redisPoolManager.
GetClient();//Obtain the connection.
                     RedisNativeClient redisNativeClient = (RedisNativ
eClient)redisClient;
                     redisNativeClient.Client = null;// KVStore for
Redis does not support the CLIENT SETNAME command. Set Client to null.
                     try
                     {
                         string key = "test-aliyun1111";
```

```
                        string value = "test-aliyun-value1111";
                        redisClient.Set(key, value);
                        string listKey = "test-aliyun-list";
                        redisClient.AddItemToList(listKey, value);
                        System.Console.WriteLine("set key " + key +
 " value " + value);
                        string getValue = redisClient.GetValue(key);
                        System.Console.WriteLine("get key " +
getValue);
                        redisClient.Dispose();//
                    }
                    catch (Exception e)
                    {
                        System.Console.WriteLine(e.Message);
                    }
                }
                System.Console.Read();
            }
            static void Main(string[] args)
            {
                //Single-connection mode
                RedisClientTest();
                //Connection-pool mode
                RedisPoolClientTest();
            }
        }
    }
```

For more information about API operations, see *https://github.com/ServiceStack/ServiceStack.Redis*.

15.2.5.1.7 node-redis client

This topic describes how to connect to a KVStore for Redis instance by using a node-redis client.

1. **Install a node-redis client.**

```
npm install hiredis redis
```

2. **Connect to a KVStore for Redis instance.**

```
 var redis = require("redis"),
    client = redis.createClient({detect_buffers: true});
    client.auth("password", redis.print)
```

3. **Use the KVStore for Redis instance.**

```
// Write data to the instance.
    client.set("key", "OK");
    // Query data on the instance. The instance returns data of
String type.
    client.get("key", function (err, reply) {
        console.log(reply.toString()); // print `OK`
    }));
    // If you specify a buffer, the instance returns a buffer.
    client.get(new Buffer("key"), function (err, reply) {
        console.log(reply.toString()); // print `<Buffer 4f 4b>`
    }));
```

```
        client.quit();
```

15.2.5.2 Use redis-cli

You can use the Redis command-line interface (redis-cli) program to connect to a KVStore for Redis instance.

Context

> **Notice:**
> **KVStore for Redis only supports connections over an internal network. Therefore, only the ECS instances that run on the same node as KVStore for Redis and that are installed with redis-cli can connect to KVStore for Redis for data operations.**

- **To use the redis-cli program to connect to the KVStore for Redis instance, run the following command:**

```
redis-cli -h Connection address of the instance -a Password
```

15.3 Instances

15.3.1 Create an instance

This topic describes how to create an instance in the KVStore for Redis console.

Prerequisites

- **You have requested an Alibaba Cloud account for logging on to the KVStore for Redis console.**
- **To use the Virtual Private Cloud (VPC) service, you must create a VPC in the same region as KVStore for Redis.**

> **Notice:**
> **You must specify the network type when you create the instance, and cannot modify the network type later.**

Procedure

1. **Log on to the** *KVStore for Redis console*.

2. **On the KVStore for Redis tab page, click Create Instance. On the Create Redis Instance page that appears, set parameters as required.**

   Before you select the VPC type, create a VPC. For more information, see the "Create a default VPC and VSwitch" topic of *VPC User Guide* **.**

   Table 15-2: KVStore for Redis parameters

Field	Parameter	Description
Region	Region	Specifies the region where the target KVStore for Redis instance is located.
	Zone	Specifies the zone where the target KVStore for Redis instance is located.
Basic Settings	Department	Specifies the department that the target KVStore for Redis instance belongs to.
	Project	Specifies the project that the target KVStore for Redis instance belongs to. **Notice:** After you specify the project, only the members that belong to this project can use this instance. For more information, see the "View project members" topic of *KVStore for Redis User Guide* .
Engine Information	Engine Version	Specifies the engine version of the target KVStore for Redis instance. You can select Redis 2.8 or Redis 4.0 as the engine version.

Field	Parameter	Description
Instance Specification	Architecture	Specifies an architecture type for the target KVStore for Redis instance. KVStore for Redis provides cluster and standard architectures. The cluster architecture supports large-capacity or high-performance requirements. Due to Redis single-thread mechanism, we recommend that you use a standard architecture if your business requires QPS performance of 100,000 or less. To require higher performance, select a cluster architecture.
	Node Type	Specifies a node type for the target KVStore for Redis instance. KVStore for Redis supports the dual-copy structure.
	Service Plan	Specifies a standard or premium plan. A premium plan supports advanced editions.
	Instance Specification	Specifies the specifications for the target KVStore for Redis instance. The maximum number of connections and maximum internal network bandwidth vary according to different instance specifications.

Field	Parameter	Description
Network	Network Type	On the Alibaba Cloud platform, a classic network and a VPC have thefollowing differences: · Classic network: cloud services in a classic network are not isolated. Unauthorized access to a cloud service is blocked only by a security group or a whitelist policy of the service. · VPC: a VPC helps you build an isolated network environment in Alibaba Cloud. You can customize the route table, CIDR blocks, and gateway of a VPC. In addition, to smoothly migrate applications to the cloud, you can use a leased line or virtual private network (VPN) to integrate an on-premises data center and cloud resources in a VPC into a virtual data center. **Note:** Before you select the VPC type, create a VPC. For more information, see the "Create a default VPC and VSwitch" topic of *VPC User Guide* .
Password	Set Password	Specifies the password for connecting to the target instance. Select Now to start setting the password, or select Later to set the password after creating the instance. For more information about how to set the password, see *Reset a password*. A password must follow these rules: · The password must be 8 to 30 characters in length. · The password must contain uppercase and lowercase letters and numbers at the same time, and does not support special characters.
Instance Name	Instance Name	Specifies the name of the target instance. · An instance name must be 2 to 128 characters in length. · The name must start with an uppercase or lowercase letter or a Chinese character and can contain letters, numbers, underscores (_), and hyphens(-).

3. **Click Create.**

   Afterward, wait until the instance stays in Normal status.

15.3.2 View details of an instance

After you create an instance, you can view details of the instance in the KVStore for Redis console.

Procedure

1. *Log on to the KVStore for Redis console*.

2. On the KVStore for Redis tab page, click the target instance ID or choose  > View Details in the Actions column next to the target instance. On the Instance Information page that appears, view details of the instance.

   The Instance Information page shows the Basic Information, Configuration Information, and Connection Information fields. These fields are described as follows:

   - Basic Information
     - Instance ID
     - Name
     - Status
     - Region
     - Department
     - Project
     - Created At
     - Zone
     - Network Type
   - Configuration Information
     - Instance Specification
     - Maximum Connections
     - Maximum Bandwidth
     - Maintenance Window
     - Whitelist
   - Connection Information
     - Address
     - Port Number
     - SSL Status (not supported by cluster instances)
     - SSL Expires At (not supported by cluster instances)

15.3.3 Modify an instance name

After you create an instance, you can change the instance name in the Apsara Stack console. You can easily locate an instance based on the instance name.

Procedure

1. *Log on to the KVStore for Redis console*.
2. On the KVStore for Redis tab page, choose  > Change Basic Information in the Actions column next to a target instance.
3. In the Change Instance Information dialog box, enter a new instance name, and click OK.

> **Note:**
> An instance name must follow these rules:
>
> - An instance name must be 2 to 128 characters in length.
> - The name must start with an uppercase or lowercase letter or a Chinese character and can contain letters, numbers, underscores (_), and hyphens(-).

15.3.4 Change the specifications

KVStore for Redis allows you to change the specifications.

Context

> **Notice:**
> When you change the specifications, a temporary disconnection may occur. We recommend that you perform this operation during off-peak hours.

Procedure

1. *Log on to the KVStore for Redis console*.
2. On the KVStore for Redis tab page, choose  > Change Specification in the Actions column next to a target instance. You can also click the target instance ID to enter the Instance information tab page, and click Change Instance.
3. In the Change Instance dialog box that appears, set Instance Specification, and then click OK.

   Afterward, the system indicates that `the instance is changed`. Wait a short period of time before the instance changes to the Normal status, and then you can use the instance.

15.3.5 Set a whitelist

To use an instance, you must specify one or more IP addresses or CIDR blocks as a whitelist. For more information, see *Set a whitelist*.

15.3.6 Set O&M time

To ensure the stability of KVStore for Redis instances, the backend system irregularly maintains instances and hosts.

Context

To guarantee the stability of the maintenance process, the instances change to the Being Maintained status before the specified operations and maintenance (O&M) time. When an instance remains in this status, you can still access data in the database as normal. You can query data as normal, such as performance monitoring data, but cannot modify the instance, for example, changing the configuration, in the console.

Notice:
During the specified O&M time, a temporary disconnection may occur. We recommend that you specify off-peak hours as the O&M time.

Procedure

1. *Log on to the KVStore for Redis console*.
2. **On the KVStore for Redis tab page, click the target instance ID or choose**

   > View Details in the Actions column next to the target instance to go to the Instance Information tab page.
3. **Click Change O&M Time.**
4. **In the Change O&M Time dialog box that appears, set Maintenance Time, and click OK.**

15.3.7 Enable SSL

To secure data on your instance, you can enable the Secure Sockets Layer (SSL)-encrypted connection after you create the instance.

Context

To enhance link security, you can enable SSL encryption and install SSL certificat e authority (CA) certificates on the required applications. SSL encrypts network connections at the transport layer. This enhances data security, but also increases connection response time.

> **Note:**
> **You cannot use SSL encryption for cluster instances.**

Procedure

1. *Log on to the KVStore for Redis console*.
2. **On the KVStore for Redis tab page, click the target instance ID or choose** 

   > View Details in the Actions column next to the target instance to go to the Instance Information tab page.
3. **Click Enable SSL. Afterward, the system indicates a successful operation. Wait a certain period and refresh the Instance Information page. You can see Disable SSL and SSL Certificate Download on the page. This indicates that SSL encryption has been enabled.**

15.3.8 Clear instance data

You can easily clear all data on an instance in the console.

Prerequisites

> **Notice:**
> **Be aware that this operation erases all data of the instance. You cannot restore the erased data.**

Procedure

1. *Log on to the KVStore for Redis console*.
2. **On the KVStore for Redis tab page, choose**  **> Clear in the Actions column next to the target instance.**
3. **On the Clear Instance dialog box that appears, click OK.**

15.3.9 Reset a password

If you forgot your password or did not configure a password when creating an instance, you can reset the password of the instance.

Procedure

1. *Log on to the KVStore for Redis console*.
2. On the KVStore for Redis tab page, click the target instance ID or choose  > View Details in the Actions column next to the target instance to go to the Instance Information tab page.
3. Click Reset Password. In the Reset Password dialog box that appears, enter the logon password and confirm the password , and then click Submit.

15.3.10 Release an instance

You can easily release a target instance in the console.

Prerequisites

Notice:
Be aware that a released instance is deleted. You cannot restore the released instance.

Procedure

1. *Log on to the KVStore for Redis console*.
2. On the KVStore for Redis tab page, choose  > Delete in the Actions column next to the target instance.
3. In the Delete Instance dialog box that appears, click OK.

15.3.11 Set parameters

KVStore for Redis allows you to set some instance parameters. For more information about the parameters that you can modify, see parameters on the Parameters page in the KVStore for Redis console.

Context

The KVStore for Redis service is fully compatible with the Redis database service. You can set parameters for KVStore for Redis and on-premises databases in a similar way. In this example, you can modify parameters in the KVStore for Redis

console, or use the Redis command-line interface (redis-cli) to run commands and set parameters.

For more information about how to set parameters for different database versions, see:

- *redis.conf for Redis 3.0*
- *redis.conf for Redis 2.8*

Procedure

1. *Log on to the KVStore for Redis console*.
2. On the KVStore for Redis tab page, click the target instance ID or choose  > View Details in the Actions column next to the target instance to go to the Instance Information tab page.
3. Click the Parameters tab.
4. Choose  > Change Basic Information in the Actions column next to the target parameter.
5. Modify the parameter and click OK.

15.4 Back up and restore data

15.4.1 Configure automatic backup policies

You can configure an automatic backup policy in the console.

Context

An increasing number of applications use KVStore for Redis for persistent storage. Therefore, KVStore for Redis supports routine backup mechanisms to easily restore data upon misoperations. Alibaba Cloud provides secondary nodes to back up .rdb files as snapshots. This backup operation does not affect the performance of your instance. You can easily customize the backup operation in the console.

 **Note:**
You cannot back up and restore data for cluster instances.

Procedure

1. *Log on to the KVStore for Redis console*.

2. On the KVStore for Redis tab page, click the target instance ID or choose 

   > View Details in the Actions column next to the target instance to go to the Instance Information tab page.
3. Click the Backup and Restore tab.
4. Click the Backup Settings tab.
5. Click Change Settings. In the Backup Settings dialog box that appears, specify the automatic backup recurrence and time.

   Notice:
   KVStore for Redis retains backup data for up to seven days by default. You cannot modify this default parameter.
6. Click OK.

15.4.2 Back up data manually

In addition to regularly scheduled backups, you can also back up data manually in the console.

Context

Note:
You cannot back up and restore data for cluster instances.

Procedure

1. *Log on to the KVStore for Redis console*.
2. On the KVStore for Redis tab page, click the target instance ID or choose 

   > View Details in the Actions column next to the target instance to go to the Instance Information tab page.
3. Click the Backup and Restore tab.
4. Click the Backups tab.
5. Click Create Backup in the upper-right corner.
6. In the Back Up Instance dialog box that appears, click OK to back up data on the instance immediately.

   Note:

On the Backups tab page, you can set Time Range to query historical backup data. KVStore for Redis retains backup data for up to seven days by default. You can query the historical backup data over the last seven days or fewer.

15.4.3 Archive backups

You can download the backup sets of an instance over the last seven days in the KVStore for Redis console.

Context

Due to industry regulations or corporate systems, you may regularly back up and archive Redis data. KVStore for Redis allows you to archive backups. This service automatically saves automatic or manual backup files to Object Storage Service (OSS). Alibaba Cloud stores your backup files in OSS for seven days. The system automatically deletes the backup files that have been stored for more than seven days.

To archive these backup files for a longer period, you can copy the link in the console and manually download the database backup files for local storage.

Note:
You cannot back up and restore data for cluster instances.

Procedure

1. *Log on to the KVStore for Redis console*.
2. On the KVStore for Redis tab page, click the target instance ID or choose 
   > View Details in the Actions column next to the target instance to go to the Instance Information tab page.
3. Click the Backup and Restore tab.
4. On the Backups tab page, click  next to the backup set that you want to download, and then select Download.
5. In the dialog box that appears, click OK to download the file to the default local directory.

15.4.4 Restore data

You can use the data recovery feature to minimize the losses caused by database misoperations. KVStore for Redis supports data backups and data recovery.

Prerequisites

Note:

- Data recovery is a risky operation. Verify that the data is correct before data recovery.
- You cannot back up and restore data for cluster instances.

You must back up data before data recovery. After the backup operation, KVStore for Redis retains backup data for up to seven days by default.

Procedure

1. *Log on to the KVStore for Redis console*.
2. On the KVStore for Redis tab page, click the target instance ID or choose  > View Details in the Actions column next to the target instance to go to the Instance Information tab page.
3. Click the Backup and Restore tab to go to the Backups tab page.
4. Set Time Range for the data you want to restore. Click Search to retrieve the backup sets within the specified period.

   Backups exist in the specified period only after you have made the backups during the period.
5. Choose  > Restore in the Actions column next to the target backup file.
6. In the Restore dialog box that appears, select OK to restore the data of the original instance.

15.5 Import data

You can use the Redis command-line interface (redis-cli) tool to import existing Redis data to KVStore for Redis based on append-only files (AOFs).

Context

The redis-cli tool is a built-in Redis CLI. KVStore for Redis allows you to use redis-cli to seamlessly import existing Redis data to KVStore for Redis.

Notes:

- KVStore for Redis only supports connections over the internal network of Apsara Stack. You must use the redis-cli tool on an ECS instance in Apsara Stack. If your Redis instance does not run on the ECS instance in Apsara Stack, you must copy the existing AOF file to the ECS instance before importing data.
- The redis-cli tool is a built-in Redis CLI. If you cannot use redis-cli on the ECS instance, you need to download and install Redis to use the tool.

If you have created a Redis instance on the ECS instance in Apsara Stack, follow these steps:

Procedure

1. Enable the AOF feature for the existing Redis instance. Skip this step if you have enabled this feature for the instance.

```
# redis-cli -h old_instance_ip -p old_instance_port config set
 appendonly yes
```

2. Use the AOF file to import data to the new KVStore for Redis instance. In this example, the AOF file is named as appendonly.aof.

```
# redis-cli -h aliyun_redis_instance_ip -p 6379 -a password --pipe <
 appendonly.aof
```

Notice:

If the AOF feature is not required for the original Redis instance, run the following command to disable the AOF feature after importing data:

```
# redis-cli -h old_instance_ip -p old_instance_port config set
 appendonly no
```

15.6 Supported Redis commands

KVStore for Redis is compatible with Redis 3.0 and supports Redis 3.0 GEO commands. For more information about Redis commands, see *http://redis.io/commands*.

Supported Redis commands

Key	String	Hash	List	Set	SortedSet
DEL	APPEND	HDEL	BLPOP	SADD	ZADD
DUMP	BITCOUNT	HEXISTS	BRPOP	SCARD	ZCARD
EXISTS	BITOP	HGET	BRPOPLPUSH	SDIFF	ZCOUNT
EXPIRE	BITPOS	HGETALL	LINDEX	SDIFFSTORE	ZINCRBY
EXPIREAT	DECR	HINCRBY	LINSERT	SINTER	ZRANGE
MOVE	DECRBY	HINCRBYFLOAT	LLEN	SINTERSTORE	ZRANGEBYSCORE
PERSIST	GET	HKEYS	LPOP	SISMEMBER	ZRANK
PEXPIRE	GETBIT	HLEN	LPUSH	SMEMBERS	ZREM
PEXPTREAT	GETRANGE	HMGET	LPUSHX	SMOVE	ZREMRANGEBYRANK
PTTL	GETSET	HMSET	LRANGE	SPOP	ZREMRANGEBYSCORE
RANDOMKEY	INCR	HSET	LREM	SRANDMEMBER	ZREVRANGE
RENAME	INCRBY	HSETNX	LSET	SREM	ZREVRANGEBYSCORE
RENAMENX	INCRBYFLOAT	HVALS	LTRIM	SUNION	ZREVRANK
RESTORE	MGET	HSCAN	RPOP	SUNIONSTORE	ZSCORE
SORT	MSET	-	RPOPLPUSH	SSCAN	ZUNIONSTORE

Key	String	Hash	List	Set	SortedSet
TTL	MSETNX	-	RPUSH	-	ZINTERSTORE
TYPE	PSETEX	-	RPUSHX	-	ZSCAN
SCAN	SET	-	-	-	ZRANGEBYLEX
OBJECT	SETBIT	-	-	-	ZLEXCOUNT
-	SETEX	-	-	-	ZREMRANGEBYLEX
-	SETNX	-	-	-	-
-	SETRANGE	-	-	-	-
-	STRLEN	-	-	-	-

HyperLogLog	Pub/Sub	Transaction	Connection	Server	Scripting	Geo
PFADD	PSUBSCRIBE	DISCARD	AUTH	FLUSHALL	EVAL	GEOADD
PFCOUNT	PUBLISH	EXEC	ECHO	FLUSHDB	EVALSHA	GEOHASH
PFMERGE	PUBSUB	MULTI	PING	DBSIZE	SCRIPT EXISTS	GEOPOS
-	PUNSUBSCRIBE	UNWATCH	QUIT	TIME	SCRIPT FLUSH	GEODIST
-	SUBSCRIBE	WATCH	SELECT	INFO	SCRIPT KILL	GEORADIUS
-	UNSUBSCRIBE	-	-	KEYS	SCRIPT LOAD	GEORADIUSBYMEMBER
-	-	-	-	CLIENT KILL	-	-
-	-	-	-	CLIENT LIST	-	-
-	-	-	-	CLIENT GETNAME	-	-
-	-	-	-	CLIENT SETNAME	-	-

HyperLogLog	Pub/Sub	Transaction	Connection	Server	Scripting	Geo
-	-	-	-	CONFIG GET	-	-
-	-	-	-	MONITOR	-	-
-	-	-	-	SLOWLOG	-	-

Unsupported commands

Key	Server
MIGRATE	BGREWRITEAOF
-	BGSAVE
-	CONFIG REWRITE
-	CONFIG SET
-	CONFIG RESETSTAT
-	COMMAND
-	COMMAND COUNT
-	COMMAND GETKEYS
-	COMMAND INFO
-	DEBUG OBJECT
-	DEBUG SEGFAULT
-	LASTSAVE
-	ROLE
-	SAVE
-	SHUTDOWN
-	SLAVEOF
-	SYNC

Commands unsupported by cluster instances

Transaction	Scripting	Connection	Key	List
DISCARD	EVAL	SELECT	MOVE	BLPOP
EXEC	EVALSHA	-	SCAN	BRPOP
MULTI	SCRIPT EXISTS	-	-	BRPOPLPUSH

Transaction	Scripting	Connection	Key	List
UNWATCH	SCRIPT FLUSH	-	-	-
WATCH	SCRIPT KILL	-	-	-
-	SCRIPT LOAD	-	-	-

Commands restricted for cluster instances

Key	String	List	Set	Sorted Set	HyperLogLog
RENAME	MSETNX	RPOPLPUSH	SINTERSTORE	ZUNIONSTORE	PFMERGE
RENAMENX	-	-	SINTER	ZINTERSTORE	-
-	-	-	SUNIONSTORE	-	-
-	-	-	SUNION	-	-
-	-	-	SDIFFSTORE	-	-
-	-	-	SDIFF	-	-
-	-	-	SMOVE	-	-

Note:

Restricted commands only support scenarios where target keys are distributed in a single hash slot. You cannot merge data from multiple hash slots. Therefore, you need to use hash tags to distribute all target keys to only one hash slot. For example, when you process three keys, key1, aakey, and abkey3, you need to store them as {key}1, aa{key}, and ab{key}3 to effectively call restricted commands. For more information about how to use hash tags, see the *Official Redis Documentation*.

16 Server Load Balancer (SLB)

16.1 What is Server Load Balancer?

Server Load Balancer (SLB) is a traffic distribution control service that distributes traffic to multiple backend Elastic Compute Service (ECS) instances based on routing algorithms and forwarding rules.

SLB is a complementary service for ECS multi-machine solutions, and must be used in conjunction with ECS. SLB expands application service capabilities by distributing and balancing traffic. It checks the health status of added backend servers and automatically isolates abnormal ECS instances to eliminate single points of failure (SPOFs), improving the overall service capabilities of your applications.

SLB provides the following functions:

- **Protocol support: SLB provides both Layer-4 (TCP and UDP) and Layer-7 (HTTP and HTTPS) load balancing services.**
- **Health check: SLB checks the health status of backend ECS instances. SLB can automatically block abnormal ECS instances and begin distributing requests to these ECS instances only when they become functional again.**
- **Session persistence: SLB provides the session persistence function. You can configure rules to forward a session request from a client to the same backend ECS instance during the session lifecycle.**
- **Routing algorithm: SLB supports round-robin and least-connections routing algorithms.**
  - **Round robin: External requests are sequentially distributed to backend ECS instances based on the number of visits.**
  - **Least connections: Backend ECS instances with fewer connections will be prioritized and accessed more frequently (probably).**
- **Domain name-based and URL-based forwarding: For Layer-7 (HTTP and HTTPS) protocols, SLB forwards requests to different VServer groups based on the preset domain names or URLs.**
- **Certificate management: SLB provides centralized certificate management for applications that use HTTPS. Certificates do not need to be uploaded to backend**

ECS instances. SLB performs decryption on access addresses, which reduces the CPU overheads of backend ECS instances.

16.2 Planning and preparation

Before creating an SLB instance, you must make a plan for the deployment of backend ECS instances, the SLB instance type (internal or external), and the listener protocols to be configured.

Make the following preparations before you create an SLB instance:

- **Create ECS instances**

  ECS instances will be added as backend servers to the SLB instance to receive and process requests forwarded by listeners. Before using SLB, you must create ECS instances and deploy applications on them. Make sure that the department of each ECS instance is the same as that of the SLB instance, and the security groups of the ECS instances allow HTTP or HTTPS access over port 80 or 443.

- **Plan the SLB instance type**

  There are two types of SLB instances: external SLB instance and internal SLB instance. Different IP addresses are allocated to each type of SLB instances. You can set the SLB instance type as needed.

  - **External: External SLB instances distribute only requests from the Internet . After you create an external SLB instance, the system will allocate a public IP address to the instance. You can associate a domain name to the public IP address to provide external services.**
  - **Internal: Internal SLB instances distribute only requests from the intranet. When configuring an internal SLB instance, you must set Network Type to Classic Network or VPC:**

    - **If Network Type of an internal SLB instance is set to Classic Network, the IP address of the SLB instance is allocated and managed by Apsara Stack in a unified manner. The SLB instance is accessible to the classic-network ECS instances that belong to the same region as the SLB instance.**
    - **If Network Type of an internal SLB instance is set to VPC, the IP address of the SLB instance is allocated from the CIDR block of the specified VSwitch. The SLB instance is accessible only to the ECS instances that belong to the same VPC as the SLB instance.**

- **Plan listeners**

  Alibaba Cloud Server Load Balancer supports both Layer-4 (TCP and UDP) and Layer-7 (HTTP and HTTPS) load-balancing services. You can configure different listeners as needed.

  Compared with Layer-4 listeners, Layer-7 listeners require an additional step of Tengine processing. Therefore, the performance of Layer-7 listeners is inferior to that of Layer-4 listeners. In addition, Layer-7 listener performance may be further deteriorated by factors such as an insufficient number of client ports and too many backend server connections. Layer-4 listeners are recommended for high performance purposes.

16.3 Quick start

16.3.1 Overview

This topic describes how to quickly create an external SLB instance and forward client requests to two backend ECS instances.

For this example, two ECS instances where the Apache Web application is built are added as backend servers to the SLB instance, to receive requests forwarded by the listeners.

Note:
To create multiple listeners to forward requests to different ECS instances, you must create VServer groups. For more information about how to create a VServer group, see *Add a VServer group*.

The following operations are involved:

- *Create an SLB instance*

  An SLB instance is a running entity of the SLB service. You can add multiple listeners and backend servers to an SLB instance.

- *Add a listener*

  You must add at least one listener to an SLB instance to forward requests to the backend servers. When configuring a listener, you need to set its basic forwarding rules and health check parameters.

- *Add backend servers*

  After configuring listeners, you must add backend servers to the SLB instance to receive and process requests forwarded by the listeners.

16.3.2 Log on to the SLB console

This topic describes how to use Chrome to log on to the SLB console from Apsara Stack Console.

Prerequisites

- **Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://`IP address or domain name of the Apsara Stack console`/manage.**
- **We recommend that you use the Chrome browser.**

Procedure

1. **Open your browser.**
2. **In the address bar, enter the access address of the Apsara Stack console in the format of http://`IP address or domain name of the Apsara Stack console`/manage, and then press Enter.**
3. **Enter the correct username and password.**
   - **The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.**
   - **You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).**
4. **Click LOGIN to go to the Dashboard page.**
5. **In the top navigation bar, choose  > Compute, Storage & Networking > Server Load Balancer.**

16.3.3 Create an SLB instance

An SLB instance is a running entity of the SLB service. You can add multiple listeners and backend servers to an SLB instance.

Prerequisites

- **You have created ECS instances and deployed applications on them.**
- **Make sure that the department of each ECS instance is the same as that of the SLB instance, and the security groups of the ECS instances allow HTTP or HTTPS access over port 80 or 443.**

Procedure

1. *Log on to the SLB console*.
2. **On the Instances page, click Create Instance.**
   - **Department: Select a department for the SLB instance from the drop-down list.**

     Note:
     Make sure that the department of the SLB instance is the same as that of the backend ECS instances.
   - **Project: Select a project for the SLB instance.**
   - **Name: Enter a name for the SLB instance.**
   - **Network Type: Set the instance type and network type of the SLB instance. For this example, select External, and leave IP Address empty. The IP address allocated by the system is used.**
3. **Click Create.**

What's next

Add a listener

16.3.4 Add a listener

You must add at least one listener to an SLB instance to forward the requests from clients to backend servers. When configuring a listener, you need to set its basic forwarding rules and health check parameters.

Prerequisites

Create an SLB instance

Procedure

1. *Log on to the SLB console*.
2. On the Instances page, click the ID of the target SLB instance.
3. On the SLB Instance page, click the Listeners tab.
4. On the Listeners tab, click Add.
5. In the Add Listener dialog box that appears, configure a listener.

   In this example, the listener is configured as follows:

   - SLB Protocol [Port]: HTTP 80
   - Backend Protocol [Port]: HTTP 80
   - Routing Algorithms: Round Robin
   - Peak Bandwidth: 100 Mbit/s
   - Session Persistence: disabled
   - Health Check Settings: default settings
6. Click OK.

What's next

Add backend servers

16.3.5 Add backend servers

After configuring listeners, you must add backend servers to the SLB instance to receive and process requests forwarded by the listeners.

Procedure

1. *Log on to the SLB console*.
2. On the Instances page, click the ID of the target SLB instance.
3. On the SLB Instance page, click the Backend Servers tab.
4. Click Add Backend Server.
5. In the dialog box that appears, select ECS instances and click Add.

   An ECS instance with a higher weight receives a greater proportion of access requests. You can set the weights of backend ECS instances based on their service capabilities. The default weight is used in this example.

16.4 SLB instances

16.4.1 SLB instance overview

An SLB instance is a running entity of the SLB service. To use the SLB service, you must create an SLB instance and add listeners and backend servers to the instance.

In the SLB console, you can edit, delete, start, and stop an SLB instance.

16.4.2 Create an SLB instance

Before using SLB, you must create an SLB instance. You can add multiple listeners and backend servers to an SLB instance.

Procedure

1. *Log on to the SLB console*.
2. **On the Instances tab page, click Create Instance.**
3. **On the Create SLB Instance page, configure the SLB instance.**

Table 16-1: Parameters for creating an SLB instance

Parameter	Description
Region	The region to which the SLB instance to be created belongs.
Zone	The zone within the selected region, to which the SLB instance belongs. If Apsara Stack is deployed in two data centers, select both a primary zone and a secondary zone for the SLB instance.
Department	The department to which the SLB instance belongs.
Project	The project to which the SLB instance belongs.
Name	The name of the SLB instance. The name must be 1 to 63 characters in length. It must start with a letter and can contain numbers, letters, hyphens (-), and underscores (_).

Parameter	Description
Instance Type	The type of the SLB instance. · Internal: If your SLB instance distributes only internal traffic, set Instance Type to Internal. For internal SLB instances, you must set the network type to Classic Network or VPC. If you create the SLB instance in a VPC, select a VPC and a VSwitch for the SLB instance. · External: If your SLB instance needs to distribute external traffic, set Instance Type to External.
Network Type	The network type of the SLB instance. · Classic Network: The instance IP address is allocated by the Apsara Stack platform. · VPC: The instance IP address is allocated by the specified VSwitch.
IP Address	The service IP address of an SLB instance. If the service IP address is left empty, the system automatically allocates an IP address to the instance based on the instance network type.
Instances	The number of instances to be created. The system creates SLB instances with the same configurations in batches.

4. Click Create.

16.4.3 Start or stop an instance

You can start or stop an SLB instance as needed.

Procedure

1. *Log on to the SLB console*.
2. On the Instances tab page, locate the target instance. In the Actions column, click the  icon and then click Start or Stop.
3. In the displayed dialog box, click OK.

16.4.4 View instance details

You can view the details such as departments, projects, and IP addresses of SLB instances in the Server Load Balancer console.

Procedure

1. *Log on to the SLB console*.
2. **On the Instances tab page, locate the SLB instance that you want to view.**
3. **Click the instance ID, or click the**  **icon in the Actions column and choose View Details from the shortcut menu, to go to the instance details page.**

16.4.5 Modify attributes of an SLB instance

You can edit the name and description of an SLB instance.

Procedure

1. *Log on to the SLB console*.
2. **On the Instances tab page, locate the target SLB instance. Click the**  **icon in the Actions column and choose Change from the shortcut menu.**
3. **In the displayed Change SLB Instance dialog box, change the name and description of the SLB instance.**
4. **Click OK.**

16.4.6 Modify the department and project of an SLB instance

You can modify the department and project of an SLB instance.

Procedure

1. *Log on to the SLB console*.
2. **On the Instances tab page, locate the target instance. Click the**  **icon in the Actions column and choose Change Ownership from the shortcut menu.**
3. **In the displayed Change Ownership dialog box, change the department and project of the instance.**
4. **Click OK.**

16.4.7 Delete an SLB instance

You can delete an SLB instance.

Procedure

1. *Log on to the SLB console*.
2. **On the Instances tab page, locate the target instance. Click the**  **icon in the Actions column and choose Delete Instance from the shortcut menu.**
3. **In the displayed dialog box, click OK.**

16.5 Listeners

16.5.1 Listener overview

SLB listeners monitor requests received by SLB instances and forward the requests to backend ECS instances based on the forwarding rules.

SLB supports Layer-4 (TCP and UDP) and Layer-7 (HTTP and HTTPS) listeners. The following table describes the features and usage scenarios of each protocol.

Table 16-2: Protocols of SLB listeners

Protocol	Feature	Scenario
TCP	· A connection-oriented protocol. A reliable connection must be established with the peer before data can be sent and received. · Source address-based session persistence. · Source address available at the network layer. · Fast data transmission.	TCP is applicable to scenarios with high requirements on reliability and data accuracy but with tolerance for low speeds, such as file transmission, email sending or receiving, remote logon, and Web applications with no special requirements.
UDP	· A non-connection-oriented protocol. Before sending data, UDP directly transmits data packets without making three-way handshake with the peer, and provides no error recovery or data retransmission. · Fast data transmission and low reliability.	UDP is applicable to scenarios where real-time transmission is more important than reliability, such as video chat and real-time financial market pushes.

Protocol	Feature	Scenario
HTTP	· An application-layer protocol primarily used to package data. · Cookie-based session persistence. · Get the source address by using X-Forwarded-For.	HTTP is applicable to applications that need to identify data content, such as Web applications and small-size mobile games.
HTTPS	· Similar to HTTP, but with an encrypted connection that prevents unauthorized access. · Centralized certificate management service. You can upload certificates to SLB. The decryption operations are completed directly on SLB.	HTTPS is applicable to applications that require encrypted transmission.

16.5.2 Configure a Layer-4 listener

Alibaba Cloud provides Layer-4 (TCP and UDP) load balancing services. Layer-4 SLB listeners forward requests directly to backend ECS instances without modifying the packet headers.

Context

For more information about the features and usage scenarios of UDP and TCP, see *Listeners*.

Procedure

1. *Log on to the SLB console*.
2. **Click the ID of the target SLB instance to go to the instance details page. Then click the Listeners tab.**
3. **On the Listeners tab, click Add.**

4. **In the Add Listener dialog box that appears, configure a Layer-4 listener.**

Table 16-3: Parameters for configuring a Layer-4 listener

Area	Parameter	Description
Basic Settings	**SLB Protocol [Port]**	**The SLB frontend protocol and port that are used to receive requests and forward requests to backend servers.** **To configure a Layer-4 listener, select TCP or UDP from the drop-down list.**
	Backend Protocol [Port]	**The port of applications deployed on backend ECS instances.**
	Scheduling Algorithm	**The scheduling algorithm. Valid values:** · **Round Robin: Requests are sequentially distributed to backend ECS instances based on the number of visits.** · **Least Connections: Requests are forwarded to the backend ECS instance with the fewest connections.**
	Peak Bandwidth	**The bandwidth peak value for the listener, in Mbit/s.** **The minimum value is 1. The maximum value cannot exceed the bandwidth value of the SLB instance.**

Area	Parameter	Description
	Session Persistence	Indicates whether to enable session persistence. For Layer-4 (TCP and UDP) listeners, SLB supports IP address-based session persistence. SLB forwards access requests from the same IP address to the same backend ECS instance for processing. If you turn on the Session Persistence switch, you must specify the timeout period of the session in Timeout.
	Timeout	The connection timeout period.
	Idle Connection Timeout	The idle connection timeout period, in seconds. Value range: 10 to 900. If no request is received during the specified timeout period, SLB closes the connection and starts a new connection when the next request comes.

Area	Parameter	Description
	Enable VServer Group	Indicates whether to use a VServer group. If you turn on the Enable VServer Group switch, select a VServer group to bind to the listener. A VServer group consists of multiple ECS instances that provide the same services. Client requests are forwarded to the ECS instances in the specified VServer group based on the forwarding rules configured for the listener. If no VServer group is used, client requests are forwarded to the backend ECS instances of the SLB instance based on the forwarding rules configured for the listener. Note: A VServer group cannot be changed once it is bound to a listener.
Health Check Settings	Port	The port used by the health check service to access backend ECS instances. The backend port configured for the listener is used by default.
	Response Timeout (Seconds)	The maximum response time for a health check before timing out. If a backend ECS instance does not respond to the health check requests within the specified period, the health check fails.
	Health Check Interval (Seconds)	The time interval between two consecutive health checks. All nodes in the LVS cluster perform regular health checks at the specified interval independently and in parallel on backend ECS instances.

Area	Parameter	Description
	Unhealthy Threshold	The number of consecutive failed health checks that must occur on an ECS instance for an LVS node to declare this ECS instance unhealthy.
	Healthy Threshold	The number of consecutive successful health checks that must occur on an ECS instance for an LVS node to declare this ECS instance healthy.
	Layer-4 listeners support HTTP health check.	
	Do you want to enable the Layer-7 health check ?	If the configured listening protocol is TCP, Layer-7 health check can be enabled.
	Domain Name and Health Check URI	The domain name and URl for health check. By default, SLB uses the internal IP address of a backend ECS instance to initiate an HTTP head request to the default homepage of the application server for health check. · If the page used for health check is not the default homepage of the application server, you must specify the domain name and URI for health check. · If you have defined the host field parameters for the HTTP head requests, you only need to specify the URI for health check.
	Health Status	The HTTP status codes for health check.

5. Click OK.

16.5.3 Configure a Layer-7 listener

Layer-7 (HTTP or HTTPS) SLB listening is a way to implement reverse proxy. After HTTP requests arrive at an SLB listener, the SLB instance establishes TCP connections with its backend servers. Then, the SLB instance sends the HTTP

requests to the backend servers over the new TCP connections by using HTTP, instead of forwarding the packets directly to the backend servers.

Context

For more information about the features and usage scenarios of HTTP and HTTPS, see *Listeners*.

Procedure

1. *Log on to the SLB console*.
2. Click the ID of the target SLB instance to go to the instance details page.
3. Click the Listeners tab.
4. On the Listeners tab, click Add.
5. In the Add Listener dialog box that appears, configure a Layer-7 listener.

Table 16-4: Parameters for configuring a Layer-7 listener

Area	Parameter	Description
Basic Settings	SLB Protocol [Port]	The SLB frontend protocol and port that are used to receive requests and forward requests to backend servers. To configure a Layer-7 listener, select HTTP or HTTPS from the drop-down list.
	Backend Protocol [Port]	The port of applications deployed on backend ECS instances.
	Scheduling Algorithm	The scheduling algorithm. Valid values: · Round Robin: Requests are sequentially distributed to backend ECS instances based on the number of visits. · Least Connections: Requests are forwarded to the backend ECS instance with the fewest connections.

Area	Parameter	Description
	Two-way Authentication	Indicates whether to enable two-way authentication. After two-way authentication is enabled, you must upload both the server and CA certificates. Two-way authentication is disabled and one-way authentication is enabled by default. Click Upload Certificate to upload the server and CA certificates. For more information, see *Upload a certificate*. **Note:** This option is applicable only to HTTPS listeners.
	Select Server Certificate	The server certificate. The server certificate allows your browser to verify whether the server-sent certificate is signed and issued by a trusted center. **Note:** This option is applicable only to HTTPS listeners.
	Select CA Certificate	The CA certificate. The CA certificate allows a server to verify whether a client certificate sent by your browser is trusted. If the verification fails, the connection request is denied. **Note:** This option is applicable only to HTTPS listeners with two-way authentication.

Area	Parameter	Description
	Peak Bandwidth	The bandwidth peak value for the listener, in Mbit/s. The minimum value is 1. The maximum value cannot exceed the bandwidth value of the SLB instance.
	Session Persistence	Indicates whether to enable session persistence. For Layer-7 (HTTP and HTTPS) listeners, SLB supports cookie-based session persistence.

Area	Parameter	Description
	Cookie Persistence	The cookie processing method. Valid values: · Cookie Insert: SLB adds a cookie to the first response from a backend server (inserting SERVERID to the HTTP or HTTPS response), and records the backend server. The next time the client carries this cookie to access SLB, the listener forwards the request to the recorded backend server. If you use this method, you must specify the cookie timeout period in Timeout. · Cookie Rewrite: When SLB discovers that a cookie is customized, it overwrites the original cookie in the response from a backend server with the new cookie, and records the backend server. The next time the client carries the new cookie to access SLB, the listener forwards the request to the recorded backend server. If you use this method, you must customize the cookie to be inserted in the HTTPS or HTTP response in Cookie Name and maintain the cookie timeout period in the backend ECS instances. Note: You must set the cookie processing method only when you have session persistence enabled.
	Timeout	The connection timeout period.
	Idle Connection Timeout	The idle connection timeout period, in seconds. Value range: 1 to 60. If no request is received during the specified timeout period, SLB closes the connection and starts a new connection when the next request comes.

Area	Parameter	Description
	Enable VServer Group	Indicates whether to use a VServer group. If you turn on the Enable VServer Group switch, select a VServer group to bind to the listener. A VServer group consists of multiple ECS instances that provide the same services. Client requests are forwarded to the ECS instances in the specified VServer group based on the forwarding rules configured for the listener. If no VServer group is used, client requests are forwarded to the backend ECS instances of the SLB instance based on the forwarding rules configured for the listener. Note: A VServer group cannot be changed once it is bound to a listener.
Health Check Settings	Enable Health Check	Indicates whether to enable health check. Health check is enabled by default. To ensure service availability, we recommend that you enable health check.
	Domain Name and Health Check URI	The domain name and URl for health check. By default, SLB uses the internal IP address of a backend ECS instance to initiate an HTTP head request to the default homepage of the application server for health check. · If the page used for health check is not the default homepage of the application server, you must specify the domain name and URI for health check. · If you have defined the host field parameters for the HTTP head requests, you only need to specify the URI for health check.
	Health Status	The HTTP status codes for health check.

Area	Parameter	Description
	Port	The port used by the health check service to access backend ECS instances. The backend port configured for the listener is used by default.
	Response Timeout (Seconds)	The maximum response time for a health check before timing out. If a backend ECS instance does not respond to the health check requests within the specified period, the health check fails.
	Health Check Interval (Seconds)	The time interval between two consecutive health checks. All nodes in the LVS cluster perform regular health checks at the specified interval independently and in parallel on backend ECS instances.
	Unhealthy Threshold	The number of consecutive failed health checks that must occur on an ECS instance for an LVS node to declare this ECS instance unhealthy.
	Healthy Threshold	The number of consecutive successful health checks that must occur on an ECS instance for an LVS node to declare this ECS instance healthy.

6. Click OK.

16.5.4 Configure forwarding rules

You can configure domain name-based or URL-based forwarding rules for an SLB instance that has Layer-7 listeners to distribute requests with different domain names or URLs to different ECS instances.

Context

You can add multiple forwarding rules under a single listener. Each forwarding rule is associated with a different VServer group. A VServer group consists of multiple ECS instances. For example, you can forward all read requests to one

VServer group and all write requests to another VServer group to optimize resource usage.

SLB has the following judgment rules for request forwarding:

- If a request matches a domain name-based or URL-based forwarding rule configured for a listener, the request is forward to the VServer group based on the rule.
- If a request does not match any domain name-based or URL-based forwarding rules configured for a listener to which a VServer group is bound, the request is forwarded to the VServer group.
- If none of the preceding conditions are met, the requests are forwarded to the backend ECS instances of the SLB instance based on the listener configuration.

Procedure

1. *Log on to the SLB console*.
2. Click the ID of the target SLB instance to go to the instance details page.
3. Click the Listeners tab.
4. On the Listeners tab, locate the target listener.

   Domain name-based or URL-based forwarding rules can be configured only for HTTP and HTTPS listeners.
5. In the Actions column, click and then click Configure Routing Algorithm.
6. In the Configure Routing Algorithm dialog box that appears, click + Add Routing Algorithm.
7. Configure forwarding rules based on the following principles:
   - Configure a domain name-based forwarding rule
     - When configuring a domain name-based forwarding rule, leave the URL field empty (no forward slash is required). The domain name can contain only letters, numbers, hyphens (-), and periods (.).
     - Domain names support both exact match and wildcard match. For example, www.aliyun.com is an exact domain name, whereas *.aliyun.com and *.market.aliyun.com are wildcard domain names. When a request matches multiple domain name-based forwarding rules simultaneously, an exact

match takes precedence over any wildcard match, as described in the following table.

Table 16-5: Domain name matching rule

Type	Request URL	Domain name matching rule (√ indicates that the domain name is matched, whereas x indicates that the domain name is not matched.)		
		www.aliyun.com	*.aliyun.com	*.market.aliyun.com
Exact match	www.aliyun.com	√	x	x
Wildcard match	market.aliyun.com	x	x	x
	info.market.aliyun.com	x	x	√

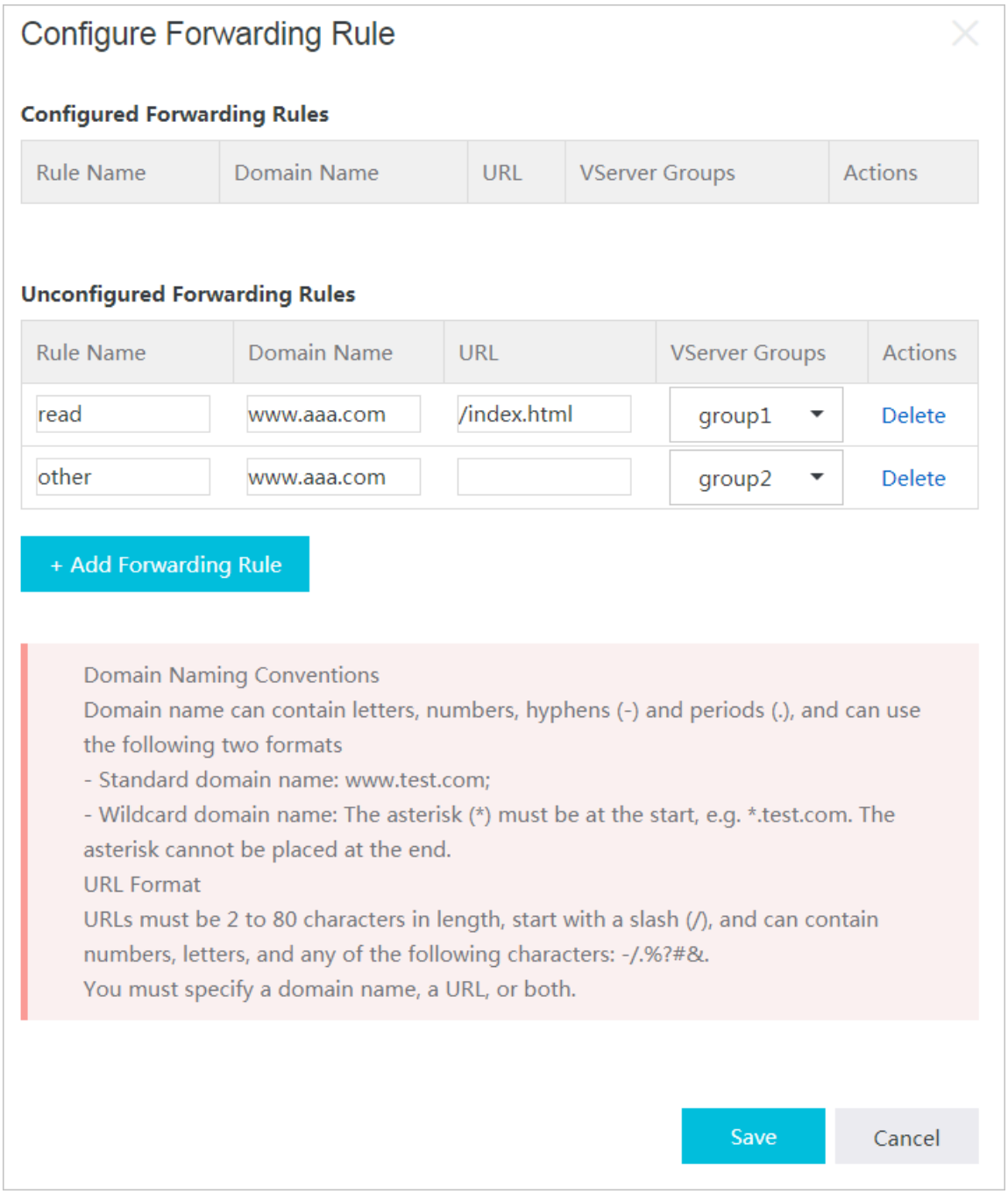
- Configure a URL-based forwarding rule
  - When configuring a URL-based forwarding rule, leave the Domain Name field empty.
  - The URL can contain only letters, numbers, hyphens (-), forward slashes (/), percent signs (%), question marks (?), number signs (#), and ampersands (&).
  - The URL must start with a forward slash (/).

    Note:
    If you enter only one forward slash (/) in the URL, the URL-based forwarding rule is invalid.
  - URL-based forwarding rules support string matching and adopt sequential matching. For example, /admin, /bbs_, and /ino_test.
- Configure a domain name- and URL-based forwarding rule

  You can combine domain name- and URL-based forwarding rules to use different URLs under the same domain name for traffic forwarding. We

recommend that you configure a default forwarding rule with the URL field left empty to prevent failures to access other unmatched URLs.

Assume that the domain name of a website is *www.aaa.com*, and that VServer Group 1 is required to process requests from *www.aaa.com/index.html* whereas VServer Group 2 is required to process other requests. To meet these processing requirements, two forwarding rules must be configured, as shown in the following figure. Otherwise, the 404 response code is returned if no forwarding rule matches the domain name *www.aaa.com*.



8. **Click Save.**

16.5.5 Configure access control

You can add a whitelist to allow specific IP addresses to access Server Load Balancer (SLB).

Context

When configuring a whitelist, note that:

- **Once a whitelist is configured, only IP addresses in the whitelist can access the SLB listener, which can lead to certain business risks.**
- **If no whitelist is configured after access control is enabled, no IP address can access the SLB listener.**
- **During the whitelist configuration process, access to the SLB listener may be interrupted for a short period of time.**

Procedure

1. *Log on to the SLB console*.
2. **Click the ID of the target SLB instance to go to the instance details page.**
3. **Click the Listeners tab.**
4. **On the Listeners tab page, locate the target listener.**
5. **In the Actions column, click the**  **icon and choose Configure Access Control from the shortcut menu.**
6. **In the Configure Access Control dialog box that appears, turn on the Enable Access Control switch.**
7. **In the Whitelist field, enter IP addresses.**

   Separate multiple IP addresses with commas (,). You can add up to 300 unique IP addresses and network segments in the form of CIDR blocks, such as 10.23.12.0/24.
8. **Click OK.**

16.5.6 Stop a listener

After a listener is stopped, it no longer forwards traffic.

Procedure

1. *Log on to the SLB console*.
2. **Click the ID of the target instance to go to the instance details page. Then click the Listeners tab.**

3. On the Listeners tab page, locate the target listener.
4. In the Actions column, click the  icon and choose Stop from the shortcut menu.

16.5.7 Start a listener

You can restart a stopped listener.

Procedure

1. *Log on to the SLB console*.
2. Click the ID of the target instance to go to the instance details page. Then, click the Listeners tab.
3. On the Listeners tab page, locate the target listener.
4. In the Actions column, click the  icon and choose Start from the shortcut menu.

16.5.8 Edit listener settings

You can edit the settings of a listener.

Procedure

1. *Log on to the SLB console*.
2. Click the ID of the target instance to go to the instance details page. Then, click the Listeners tab.
3. On the Listeners tab page, locate the target listener.
4. In the Actions column, click the  icon and choose Change from the shortcut menu.
5. Change the listener settings, and click Save.

16.5.9 Delete a listener

You can delete listeners which you no longer need.

Procedure

1. *Log on to the SLB console*.
2. Click the ID of the target instance to go to the instance details page. Then, click the Listeners tab.

3. On the Listeners tab page, locate the target listener.
4. In the Actions column, click the  icon and choose Delete from the shortcut menu.
5. In the message that appears, click OK.

16.6 Backend servers

16.6.1 Backend server overview

Before using SLB, you must add ECS instances as the backend servers of your SLB instance to receive and process requests forwarded by the listeners.

You can use SLB to virtualize multiple ECS instances in the same region into an application server pool with high performance and high availability by configurin g virtual IP addresses (VIPs). SLB performs health check on ECS instances in the application server pool, automatically blocks abnormal ECS instances, and begins distributing requests to these ECS instances only when they become functional again. The health check function improves the overall availability of services and mitigates the impact of exceptions in backend ECS instances on the services.

You can increase or decrease the number of backend ECS instances at any time. Before you perform these operations, make sure that health check is enabled and that there is at least one properly running backend ECS instance to ensure service continuity.

16.6.2 Add backend servers

After creating an SLB instance, you must add ECS instances as backend servers to your SLB instance to process the connection requests forwarded by the listeners.

Procedure

1. *Log on to the SLB console*.
2. Click the ID of the target SLB instance to go to the instance details page. Then, click the Backend Servers tab.
3. On the Backend Servers tab page, click Add Backend Server.

4. In the Add Backend Server dialog box that appears, select ECS instances and set their weights.

   An ECS instance with a higher weight receives a greater proportion of access requests. You can set the weights of ECS instances based on their service capabilities.

   > **Note:**
   > If the weight of an ECS instance is set to 0, the ECS instance no longer receives new requests.

5. Click Add.

16.6.3 Modify the weight of an ECS instance

You can modify the weight of an added ECS instance. An ECS instance with a higher weight receives a greater proportion of access requests. You can set the weights of ECS instances based on their service capabilities.

Context

> **Note:**
> If a backend ECS instance has been added to a VServer group, you must modify the weight of this ECS instance by editing the VServer group. For more information about how to edit a VServer group, see *Edit a VServer group*.

Procedure

1. *Log on to the SLB console*.
2. Click the ID of the target SLB instance to go to the instance details page. Then, click the Backend Servers tab.
3. In the Actions column corresponding to the target ECS instance, click the  icon and choose Change from the shortcut menu.
4. In the Change Backend Server dialog box that appears, change the weight of the ECS instance.

   > **Note:**
   > If the weight is set to 0, the ECS instance no longer receives new requests.

5. Click Save.

16.6.4 Remove a backend ECS instance

Directly removing an ECS instance from an SLB instance may cause intermittent service interruptions. We recommend that before removing an ECS instance from an SLB instance, you change the weight of the ECS instance to 0 so that the SLB instance no longer forwards traffic to it.

Context

> **Note:**
> **For a backend ECS instance that has been added to a VServer group, you must first remove the ECS instance from the VServer group by editing the VServer group. Then, switch to the Backend Servers tab page to change the weight of the ECS instance to 0 and remove the ECS instance. For more information about how to edit a VServer group, see *Edit a VServer group*.**

Procedure

1. *Log on to the SLB console*.
2. Click the ID of the target SLB instance to go to the instance details page. Then, click the Backend Servers tab.
3. Locate the ECS instance to be removed. In the Actions column, click the icon and choose Remove from the shortcut menu.
4. In the dialog box that appears, click OK.

16.7 VServer groups

16.7.1 Add a VServer group

A VServer group is a group of ECS instances. By using VServer groups (different listeners are associated with different VServer groups), SLB forwards requests to different ECS instances. When you need to distribute different requests to different backend servers, or when you want to configure domain name-based or URL-based forwarding rules, you can use VServer groups.

Context

VServer groups have the following restrictions:

- **An ECS instance can be added to multiple VServer groups.**
- **A VServer group can be associated with multiple listeners.**

Procedure

1. *Log on to the SLB console*.
2. Click the ID of the target SLB instance to go to the instance details page. Then click the VServer Groups tab.
3. Click Add VServer Group.
4. In the Create VServer Group dialog box, perform the following operations:
   a) Enter a name for the VServer group to be added.
   b) Select a search condition from the ECS Instance ID drop-down list, enter a value in the search box, and click Search to search for ECS instances.
   c) In the Available Servers list, click the ECS instances to be added one by one.
   d) The ECS instances are added to the Selected Servers list. Then, set the ports and weights of the added ECS instances.

      An ECS instance with a higher weight receives a greater proportion of access requests. You can set the weights of ECS instances based on their service capabilities.

      Note:
      If the weight of an ECS instance is set to 0, the ECS instance no longer receives new requests.
   e) Click OK.

16.7.2 View a VServer group

You can view information such as status and ports of the ECS instances in a VServer group.

Procedure

1. *Log on to the SLB console*.
2. Click the ID of the target SLB instance to go to the instance details page. Then click the VServer Groups tab.
3. Locate the target VServer group. Click the group ID, or click the icon in the Actions column and choose View from the shortcut menu.

16.7.3 Edit a VServer group

You can modify the name of a VServer group, modify the weights of the ECS instances in the VServer group, and add new ECS instances to or remove ECS instances from the VServer group.

Procedure

1. *Log on to the SLB console*.
2. **Click the ID of the target SLB instance to go to the instance details page. Then, click the VServer Groups tab.**
3. **In the Actions column corresponding to the target VServer group, click the**  **icon and choose Change from the shortcut menu.**
4. **Change the ECS instances in the VServer group.**
5. **Click OK.**

16.7.4 Delete a VServer group

You can delete VServer groups which you no longer need.

Prerequisites

If a VServer group to be deleted is associated with a listener, you must first delete the listener.

Procedure

1. *Log on to the SLB console*.
2. **Click the ID of the target SLB instance to go to the instance details page. Then click the VServer Groups tab.**
3. **Locate the target VServer group. In the Actions column, click the**  **icon and choose Delete from the shortcut menu.**
4. **In the dialog box that appears, click Yes.**
5. **Click OK.**

16.8 Certificates

16.8.1 Certificate overview

SLB provides a certificate management function for HTTPS listeners.

To configure HTTPS listeners, you must upload the required certificates.

- **For HTTPS two-way authentication, upload both the CA and server certificates.**
- **For HTTPS one-way authentication, upload only the server certificate.**

After the certificates are uploaded to SLB, you do not need to deploy the certificates on the backend ECS instances. Private keys uploaded to the certificate management system are encrypted and stored.

- **Server certificate: allows your browser to verify whether the server-sent certificate is signed and issued by a trusted center. You can purchase a server certificate from *Alibaba Cloud Security Certificate Service* or other service providers.**
- **Client certificate: proves your identity when you use a client to communicate with the server. You can sign a client certificate with a self-signed CA certificate.**
- **CA certificate: allows a server to verify whether a client certificate sent by your browser is trusted. If the verification fails, the connection request is denied.**

16.8.2 Certificate format

Only PEM certificates in the Linux environment can be uploaded to SLB.

Certificate format requirements

Make sure that the certificate to be uploaded meets the following requirements:

- **Certificate issued by the root CA**

  If your certificate is issued by the root CA, only this certificate is required for access devices such as browsers to trust your website. Make sure that the certificates comply with the following rules:

  - **The certificate content is placed between `-----BEGIN CERTIFICATE-----` and `-----END CERTIFICATE-----` . Ensure that the certificate includes the header and footer when you upload it.**
  - **Each row contains 64 characters, and the last row can contain fewer than 64 characters.**
  - **The certificate content cannot include any spaces.**

- **Certificate issued by an intermediate CA**

  If the certificate has been issued to you by an intermediate CA and the certificate file consists of multiple certificates, you must combine the server certificate and intermediate certificate before uploading them. Combine the certificates based on the following rules:

  - **The server certificate must be followed by the intermediate certificate. There cannot be any blank rows between the certificates.**
  - **The certificate content cannot include any spaces.**
  - **There cannot be any blank rows between the certificates. Each row contains 64 characters. For more information, see** *https://www.ietf.org/rfc/rfc1424.txt***.**
  - **The certificates must meet the format requirements. Generally, the CA provides a relevant description when issuing a certificate. Pay attention to the rule description.**

RSA private key format requirements

When you upload a server certificate, you must also upload an RSA private key for the certificate. Make sure that the RSA private key complies with the following rules:

- **The key is placed between** `-----BEGIN RSA PRIVATE KEY-----` **and** `-----END RSA PRIVATE KEY-----.` **Ensure that the key includes the header and footer when you upload it.**
- **There cannot be any blank rows in the content. Each row must contain exactly 64 characters, except for the final row. The final row can contain fewer than 64 characters.**

Note:

If your private key is encrypted (for example, the header and footer of the private key are `-----BEGIN PRIVATE KEY-----` **and** `-----END PRIVATE KEY-----` **or** `-----BEGIN ENCRYPTED PRIVATE KEY-----` **and** `-----END ENCRYPTED PRIVATE KEY-----`**) or the private key contains** `Proc-Type: 4,ENCRYPTED`**, you must first run the following command to convert the private key and upload** *new_server_key.pem* **together with the server certificate:**

```
openssl rsa -in old_server_key.pem -out new_server_key.pem
```

16.8.3 Generate a CA certificate

To configure HTTPS two-way authentication, you must generate and upload a CA certificate after purchasing a server certificate.

Context

This topic describes how to generate a self-signed CA certificate by using OpenSSL.

Procedure

1. **Run the following commands to create a** *ca* **folder under the** */root* **directory and then four subfolders under the** *ca* **folder:**

```
$ sudo mkdir ca
$ cd ca
$ sudo mkdir newcerts private conf server
```

   - *newcerts* **subfolder: is the certificate backup directory that stores the digital certificates signed by the CA.**
   - *private* **subfolder: stores the private key of the CA.**
   - *conf* **subfolder: stores the configuration files for simplifying parameters.**
   - *server* **subfolder: stores the server certificate.**

2. **Create an** *openssl.conf* **file that contains the following information under the** *conf* **directory:**

```
[ ca ]
default_ca = foo

[ foo ]
dir = /root/ca
database = /root/ca/index.txt
new_certs_dir = /root/ca/newcerts

certificate = /root/ca/private/ca.crt
serial = /root/ca/serial
private_key = /root/ca/private/ca.key
RANDFILE = /root/ca/private/.rand

default_days = 365
default_crl_days= 30
default_md = md5
unique_subject = no
policy = policy_any

[ policy_any ]
countryName = match
stateOrProvinceName = match
organizationName = match
organizationalUnitName = match
localityName = optional
commonName      = supplied
```

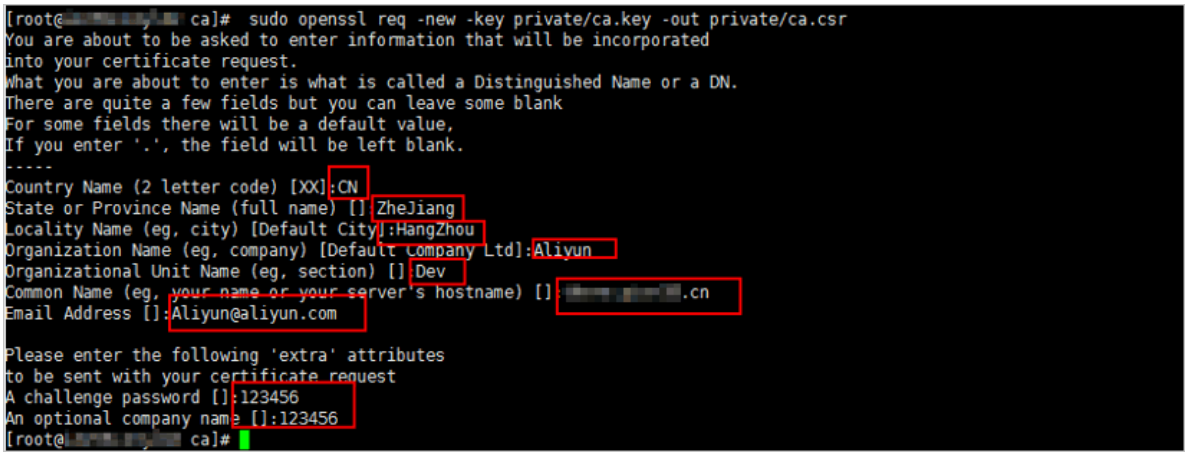
```
emailAddress     = optional
```

3. **Run the following commands to generate a private key:**

```
$ cd /root/ca
$ sudo openssl genrsa -out private/ca.key
```

4. **Run the following command, enter required information as shown in the following figure, and then press Enter to generate a certificate request *.csr* file.**

```
$ sudo openssl req -new -key private/ca.key -out private/ca.csr
```



5. **Run the following command to generate a *.crt* file:**

```
$ sudo openssl x509 -req -days 365 -in private/ca.csr -signkey
private/ca.key -out private/ca.crt
```

6. **Run the following command to set the start sequence number for the key, which can be any four characters:**

```
$ sudo echo FACE > serial
```

7. **Run the following command to create a CA key library:**

```
$ sudo touch index.txt
```

8. **Run the following command to create a certificate revocation list for removing the client certificate:**

```
$ sudo openssl ca -gencrl -out /root/ca/private/ca.crl -crldays 7 -
config "/root/ca/conf/openssl.conf"
```

The command output is as follows:

```
Using configuration from /root/ca/conf/openssl.conf
```

9. **Run the following commands to view the generated CA certificate:**

```
cd private
```

```
ls
```

16.8.4 Generate a client certificate

A client certificate proves your identity when you use a client to communicate with the server.

Prerequisites

A CA certificate is required to sign the client certificate. Make sure that you already *Generate a CA certificate*.

Procedure

1. **Run the following command to create the** `users` **directory under the** `ca` **directory to store keys:**

   ```
   $ sudo mkdir users
   ```

2. **Run the following command to create a key for the client certificate:**

   ```
   $ sudo openssl genrsa -des3 -out /root/ca/users/client.key 1024
   ```

   When creating the key, enter the pass phrase as the key password to prevent unauthorized use if the key leaks. Enter the same password twice.

3. **Run the following command to create a certificate signature request** `.csr` **file for the key:**

   ```
   $ sudo openssl req -new -key /root/ca/users/client.key -out /root/ca
   /users/client.csr
   ```

   Enter the pass phrase stored in the preceding step as prompted, press Enter, and enter the required information as prompted.

   Note:
   A challenge password is the password of the client certificate (which must be separated from the password of client.key). It can be the same as the password of the server or root certificate.

4. **Run the following command to use the CA key to sign the client key:**

   ```
   $ sudo openssl ca -in /root/ca/users/client.csr -cert /root/ca/
   private/ca.crt -keyfile /root/ca/private/ca.key -out /root/ca/users/
   client.crt -config "/root/ca/conf/openssl.conf"
   ```

   Enter `y` **twice when prompted to confirm the operation.**

5. **Run the following command to convert the certificate to a *PKCS12* file that can be recognized by most browsers:**

   ```
   $ sudo openssl pkcs12 -export -clcerts -in /root/ca/users/client.crt
    -inkey /root/ca/users/client.key -out /root/ca/users/client.p12
   ```

   Enter the pass phrase of client.key as prompted and press Enter. Then enter the password used for exporting the client certificate. This password is used to protect the client certificate, which is required when the client certificate is installed.

6. **Run the following commands to view the generated client certificate:**

   ```
   cd users
   ls
   ```

16.8.5 Upload a certificate

SLB provides the certificate management function to implement data transfer encryption and authentication over HTTPS. You can store certificates in the SLB certificate management system without deploying the certificates on backend ECS instances. Private keys uploaded to the certificate management system are encrypted and stored. A certificate can be applied to one or more listeners.

Note:
Each account can upload up to 100 certificates.

Prerequisites

You have generated a server or CA certificate to be uploaded.

Procedure

1. *Log on to the SLB console*.
2. **Click the Certificates tab to view the certificate list.**
3. **Click Upload Certificate.**

4. **In the Upload Certificate dialog box that appears, set the parameters.**

   Table 16-6: Parameters for uploading a certificate

Parameter	Description
Region	The region where a certificate is used. SLB manages certificates by regions. To use a certificate in multiple regions, upload the certificate in each region individually.
Department	The department that uses a certificate.
Project	The project that uses a certificate.
Certificate Type	The type of a certificate to be uploaded. Valid values: · CA Certificate: A CA certificate allows a server to verify whether a client certificate sent by your browser is trusted. If the verification fails, the connection request is denied. · Server Certificate: Your browser uses a server certificate to verify whether the certificate sent by the server can be trusted.
Certificate Name	The name of a certificate.
Certificate Contents	The content of a certificate. The certificate must be in the PEM format. You can click Examples to view the sample format. For more information, see *Certificate format*.
Private Key	The private key of a server certificate. Private keys must comply with the format requirements in SLB. You can click Examples to view the sample format.

5. **Click OK.**

16.8.6 Convert the format of a certificate

Server Load Balancer supports PEM certificates only. Certificates in other formats must be converted to the PEM format before they can be uploaded to Server Load Balancer.

Context

We recommend that you use OpenSSL to convert certificates. This topic describes how to convert popular certificate formats to PEM:

- **DER: This format is usually used on Java platforms.**
- **P7B: This format is usually used in Windows servers and Tomcat.**
- **PFX: This format is usually used in Windows servers.**

Convert the certificate format from DER to PEM

1. **Run the following command to convert the format of a certificate:**

```
openssl x509 -inform der -in certificate.cer -out certificate.pem
```

2. **Run the following command to convert the private key:**

```
openssl rsa -inform DER -outform PEM -in privatekey.der -out
privatekey.pem
```

Convert the certificate format from P7B to PEM

1. **Run the following command to convert the format of a certificate:**

```
openssl pkcs7 -print_certs -in incertificat.p7b -out outcertificate.
cer
```

2. **In** *outcertificat.cer*, **retrieve the** `[——-BEGIN CERTIFICATE——-, ——-END CERTIFICATE——-]` **content and upload the content as a certificate.**

Convert the certificate format from PFX to PEM

1. **Run the following command to extract the private key:**

```
openssl pkcs12 -in certname.pfx -nocerts -out key.pem -nodes
```

2. **Run the following command to extract the certificate:**

```
openssl pkcs12 -in certname.pfx -nokeys -out cert.pem
```

16.8.7 Replace a certificate

If your certificate expires or an error is reported when you upload a certificate, you can generate and upload a new certificate, and then delete the existing one.

Procedure

1. **Create and upload a new certificate.**

   For more information, see *Generate a certificate* **and** *Upload a certificate*.

2. **When you configure an HTTPS listener, you must also configure a new certificate.**

   For more information, see *Configure a Layer-7 listener*.

3. **On the Certificates page, locate the old certificate. In the Actions column, click the**  **icon and choose Delete Certificate from the shortcut menu to delete the certificate.**

17 Virtual Private Cloud (VPC)

17.1 What is VPC?

A Virtual Private Cloud (VPC) is a private network established in Apsara Stack. VPCs are logically isolated from each other.

You have full control over your VPC. For example, you can select its IP address range and configure routing tables and gateways. You can also use Alibaba Cloud resources such as ECS, RDS, and SLB in your own VPCs. You can connect a VPC to other VPCs or a local network to form an on-demand customizable network environment. This allows you to smoothly migrate applications to the cloud.

Components

Each VPC consists of a private Classless Inter-Domain Routing (CIDR) block, a VRouter, and at least a VSwitch.

- **CIDR block**

  A CIDR block is a private IP address range in a VPC. The IP addresses of all cloud resources deployed in the VPC are within the specified CIDR block. When creating a VPC or a VSwitch, you must specify the private IP address range in the form of a CIDR block.

  You can use any of the following standard CIDR blocks and their subnets as the IP address range of the VPC.

CIDR block	Number of available private IP addresses (system reserved ones excluded)
192.168.0.0/16	65,532
172.16.0.0/12	1,048,572
10.0.0.0/8	16,777,212

- **VRouter**

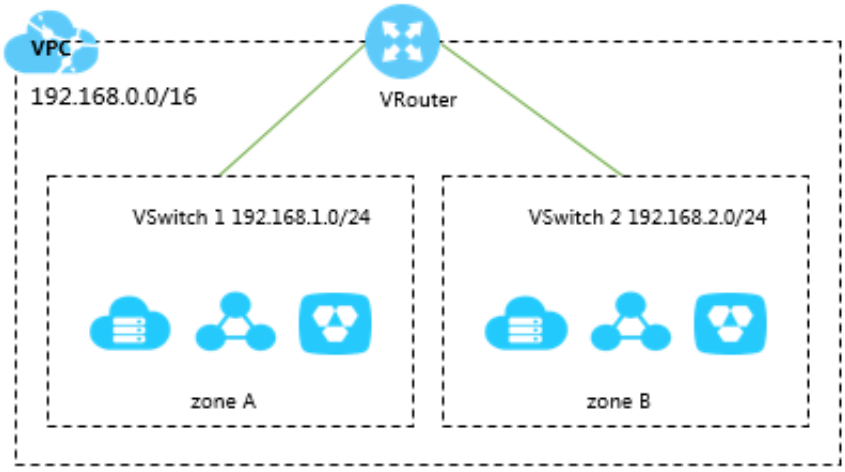
  The VRouter is the hub of a VPC. As an important component of a VPC, the VRouter connects the VSwitches in a VPC and serves as the gateway connecting

the VPC with other networks. After you create a VPC, the system automatically creates a VRouter, which is associated with a routing table.

- VSwitch

  A VSwitch is a basic network device in a VPC and is used to connect different cloud product instances. After creating a VPC, you can further divide the VPC to one or more subnets by creating VSwitches. The VSwitches within a VPC are interconnected. You can deploy applications in VSwitches of different zones to improve the service availability.

Figure 17-1: VPC



17.2 Quick start

17.2.1 Tutorial overview

This topic describes how to create a basic VPC and deploy an ECS instance in it.

Specific operations are as follows:

- *Log on to the VPC console*

  This topic describes how to log on to the VPC console.

- *Create a VPC and a VSwitch*

  You must create a VPC and VSwitch before you can deploy and use cloud services in a VPC.

- *Create a security group*

  Before creating an ECS instance in a VPC, you must first create a security group. Security groups are an important means for network security isolation. They are used to set network access control for one or more ECS instances.

- *Create an ECS instance*

  An ECS instance is a virtual computing environment that consists of most basic components of a server, such as the CPU, memory, OS, disk, and bandwidth.

17.2.2 Log on to the VPC console

This topic describes how to log on to the VPC console.

Prerequisites

- **Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://`IP address or domain name of the Apsara Stack console`/manage.**
- **We recommend that you use the Chrome browser.**

Procedure

1. **Open your browser.**
2. **In the address bar, enter the access address of the Apsara Stack console in the format of http://`IP address or domain name of the Apsara Stack console`/manage, and then press Enter.**
3. **Enter the correct username and password.**
   - **The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.**
   - **You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).**
4. **Click LOGIN to go to the Dashboard page.**

5. In the top navigation bar, choose  > Compute, Storage & Networking > Virtual Private Cloud.

17.2.3 Create a VPC and a VSwitch

You must create a VPC and a VSwitch before you can deploy and use cloud services in the VPC.

Context

When creating a VPC, note that:

- Only one CIDR block can be specified for each VPC. For more information, see *Plan CIDR blocks*.
- When a VPC is created, a VRouter and a routing table are automatically created. Each VPC can contain only one VRouter and one routing table.

Procedure

1. *Log on to the VPC console*.
2. On the VPC page, click Create.
3. Set parameters for creating a VPC. The following table describes the parameters.

Table 17-1: Parameters for creating a VPC

Parameter	Description
Name	Enter the name of the VPC. The name must be 2 to 128 characters in length. It must start with a letter. It cannot contain special characters such as at signs (@), forward slashes (/), colons (:), angle brackets (<>), curly brackets ({}), braces ([]), or spaces.
Description	Enter the description of the VPC.
Region	Select the region of the VPC.
Department	Select the department of the VPC.
Shared with Subdepartments	Specify whether to share the VPC with subdepartments. If you select Yes, subdepartment administrators can create resources in the VPC.

Parameter	Description
IP Address Range	Select the CIDR block of the VPC. Once the VPC is created, you cannot modify its CIDR block.

4. Click OK, and then click Next to create a VSwitch.
5. In the Create VSwitch dialog box that appears, set parameters for creating a VSwitch. The following table describes the parameters.

Table 17-2: Parameters for creating a VSwitch

Parameter	Description
Zone	Select the zone of the VSwitch. In a VPC, a VSwitch can be located in only one zone and cannot span across multiple zones. You can deploy cloud service instances to VSwitches in different zones to implement cross-zone disaster recovery. Note: A cloud service instance can be added to one VSwitch only.
Name	Enter the name of the VSwitch.

Parameter	Description
IP Address Range	Enter the CIDR block of the VSwitch. When setting this parameter, note that: · You must specify the IP address range for the VSwitch in the form of a CIDR block. The subnet mask of the VSwitch CIDR block can be 16 to 29 bits, which means the VSwitch can provide 8 to 65,536 IP addresses. · The CIDR block of the VSwitch must be a subset of the CIDR block of the VPC where the VSwitch resides. Note: If the CIDR block of the VSwitch is the same as that of the VPC where the VSwitch resides, you can create only one VSwitch in the VPC. · The first IP address and the last three IP addresses in each VSwitch CIDR block are reserved for system use. For example, in a VSwitch with CIDR block 192.168.1.0/24, the IP addresses 192.168.1.0, 192.168.1.253, 192.168.1.254, and 192.168.1.255 are reserved. · The CIDR block of the VSwitch cannot be the same as the destination CIDR block of a routing entry of the VPC where the VSwitch resides. However, the CIDR block of the VSwitch can be a subset of the destination CIDR block of the routing entry. · Once the VSwitch is created, you cannot modify its CIDR block.
Description	Enter the description of the VSwitch.

6. Click OK, and then click Cancel to close the dialog box.

17.2.4 Create a security group

Before creating an ECS instance in a VPC, you must create a security group. As an important means for network security isolation, security groups are used to set network access control for one or more ECS instances.

Procedure

1. Log on to Apsara Stack console.
2. In the top navigation bar, choose > Compute, Storage & Networking > Elastic Compute Service.
3. Click the Security Groups tab, and then click Create Security Group.

4. **In the Create Security Group dialog box that appears, set parameters for creating a security group. The following table describes the parameters.**

   Table 17-3: Parameters for creating a security group

Parameter	Description
Region	**Select the region of the security group.** **Make sure that the VPC and security group are in the same region.**
Department	**Select the department of the security group.** **Make sure that the VPC and security group are in the same department.**
Project	**Select the project of the security group.**
Network Type	**Select VPC.**
VPC	**Select the VPC of the security group.**
Security Group Name	**Enter the name of the security group.**
Basic Settings	**Configure the basic information of the security group.**

5. **Click OK.**

17.2.5 Create an ECS instance

An ECS instance is a virtual computing environment that consists of most basic components of a server, such as the CPU, memory, operating system, disk, and bandwidth.

Procedure

1. **Log on to Apsara Stack console.**
2. **In the top navigation bar, choose**  **> Compute, Storage & Networking > Elastic Compute Service.**
3. **Click the Instances tab, and then click Create Instance.**

4. **In the Create Instance dialog box that appears, set parameters for creating an ECS instance, and then click Create.**

   For more information about how to create an ECS instance, *see* `Create an instance` **under** `Quick start` **in the ECS user guide.**

    **Note:**
   Set Network Type to VPC, and select the VPC and VSwitch you have created.

17.3 VPC

17.3.1 Plan CIDR blocks

When creating a VPC and a VSwitch, you must specify the private IP address range for them in the form of a CIDR block.

CIDR is a bitwise, prefix-based standard for the interpretation of IP addresses. It facilitates routing by allowing blocks of addresses to be grouped into single routing table entries. You can flexibly allocate IP address segments with subnet masks such as /25, /26, and /27. These IP address segments are called CIDR blocks.

Plan the CIDR block for a VPC

When planning the CIDR block for a VPC, note that:

- **You can use the standard private network segments (192.168.0.0/16, 10.0.0.0/0, and 172.16.0.0/12) and their subnets as the CIDR block for the VPC. Only one CIDR block can be specified for each VPC. The available CIDR blocks are specified by the** `vpc_customer_private_cidr` **parameter in the global configuration during the delivery planning phase when you deploy Apsara Stack.**
- **For a VPC created by using the API, the subnet mask of the VPC can be 8 to 24 bits .**
- **Once the VPC is created, you cannot modify its CIDR block.**

Plan the CIDR block for a VSwitch

When planning the CIDR block for a VSwitch, note that:

- **The subnet mask of the VSwitch CIDR block can be 16 to 29 bits, which means the VSwitch can provide 8 to 65,536 IP addresses.**

- The CIDR block of the VSwitch must be a subset of the CIDR block of the VPC where the VSwitch resides.

  > **Note:**
  > If the CIDR block of the VSwitch is the same as that of the VPC where the VSwitch resides, you can create only one VSwitch in the VPC.

- The first IP address and the last three IP addresses in each VSwitch CIDR block are reserved for system use. For example, in a VSwitch with CIDR block 192.168.1.0/24, the IP addresses 192.168.1.0, 192.168.1.253, 192.168.1.254, and 192.168.1.255 are reserved.
- The CIDR block of the VSwitch cannot be the same as the destination CIDR block of a routing entry of the VPC where the VSwitch resides. However, the CIDR block of the VSwitch can be a subset of the destination CIDR block of the routing entry.
- Once the VSwitch is created, you cannot modify its CIDR block.

17.3.2 Create a VPC

A VPC is an isolated virtual network environment that is built in Alibaba Cloud. You have full control over your VPC. For example, you can specify its IP address range, and configure routing tables and gateways. You can also use Apsara Stack resources such as ECS, RDS, and SLB in your own VPC. You must create a VPC and VSwitch before you can deploy and use cloud services in the VPC.

Context

When creating a VPC, note that:

- Only one CIDR block can be specified for each VPC. For more information, see *Plan CIDR blocks*.
- After a VPC is created, a VRouter and a routing table are automatically created. Each VPC can contain only one VRouter and one routing table.

Procedure

1. *Log on to the VPC console*.
2. On the VPC tab page, click Create.

3. **Configure the VPC based on the following information.**

   Table 17-4: VPC configurations

Item	Configuration method
Name	Enter the name of the VPC. The name must be 2 to 128 characters in length. It must start with a letter and can contain letters, digits, periods (.), underscores (_), and hyphens (-).
Description	Enter the description of the VPC.
Region	Select a region for the VPC.
Department	Select a department for the VPC.
Shared with Subdepartments	Specify whether to allow lower-level department administrators to share VPC resources.
CIDR Block	Select a CIDR block for the VPC. After a VPC is created, its CIDR block cannot be modified. **Note:** You can set a custom CIDR block. The custom CIDR block must be the default CIDR Block in the current region or its subnet segment. The subnet mask of the CIDR block must be from 8 to 28 bits.

4. **Click OK. In the Create VPC message that appears, click Next to create a VSwitch.**

   For more information, see *Create a VSwitch*.

17.3.3 View a VPC

On the VPC tab page, you can specify filtering conditions to search for the VPC you want to view.

Procedure

1. *Log on to the VPC console*.
2. **Locate the relevant VPC, and click the instance ID or click the** icon **in the Actions column. Then choose Details from the shortcut menu to view VSwitches and VRouters in the VPC.**

17.3.4 Modify VPC information

After creating a VPC, you can modify the name and description of this VPC.

Procedure

1. *Log on to the VPC console*.
2. Locate the VPC that you want to delete, click the icon in the Actions column, and choose Edit from the shortcut menu.
3. In the Modify VPC dialog box that appears, set the name and description, and click OK.

17.3.5 Delete a VPC

You can delete a VPC if you no longer need it.

Prerequisites

Before deleting a VPC, you must release or move all resources, including VSwitches , from the VPC.

After the VPC is deleted:

- The security groups in this VPC are deleted as well.
- Data stored in this VPC cannot be restored.

Procedure

1. *Log on to the VPC console*.
2. Locate the VPC that you want to delete, click the icon in the Actions column, and choose Delete from the shortcut menu.
3. In the message that appears, click OK.

17.4 VSwitch

17.4.1 Create a VSwitch

A VSwitch is a basic network device in a VPC. It connects different cloud service instances. After a VPC is created, you can divide the VPC into several subnets by adding VSwitches.

Context

When creating a VSwitch, note that:

- You can create a maximum of 24 VSwitches for each VPC.

- After a VSwitch is created, the system automatically adds a system routing entry that is in the same CIDR block as the VSwitch.
- In a VPC, a VSwitch can be located in only one zone and cannot span across several zones. You can deploy cloud service instances on different VSwitches to implement cross-zone disaster recovery.

> **Note:**
> A VSwitch does not support multicasting or broadcasting.

Procedure

1. *Log on to the VPC console*.
2. On the VPC tab page, locate a relevant VPC and click the VPC ID.
3. Click the VSwitches tab, and click Create.
4. In the Create VSwitch dialog box that appears, configure the VSwitch based on the following information.

Table 17-5: VSwitch configurations

Item	Configuration method
Zone	Select a zone for the VSwitch. In a VPC, a VSwitch can be located in only one zone and cannot span across several zones. You can deploy cloud service instances on VSwitches in different zones to implement cross-zone disaster recovery. **Note:** A cloud service instance can be added to one VSwitch only.
Name	Enter the name of the VSwitch. The name must be 2 to 128 characters in length. It must start with a letter and can contain letters, digits, periods (.), underscores (_), and hyphens (-).

Item	Configuration method
CIDR Block	Enter the CIDR block of the VSwitch. · You must specify the network segment of a VSwitch in the form of a CIDR block. The mask length of the VSwitch CIDR block can be between 16 to 29 bits, which can provide 8 to 65,536 IP addresses. · The CIDR block of a VSwitch must be a subset of the CIDR block of the VPC. Note: If the CIDR blocks of your VSwitch and VPC are the same, only this single VSwitch can be created. · The first IP address and the last three IP addresses of each VSwitch are reserved for the system. For example, if 192.168.1.0/24 is the CIDR block of the VSwitch, 192.168.1.0, 192.168.1.253, 192.168.1.254, and 192.168.1.255 are reserved for the system. · The CIDR block of a VSwitch cannot be the same as the destination CIDR block in the routing entry of the VPC where the VSwitch resides. However, it can be a subset of the destination CIDR block in the current routing entry. · After a VSwitch is created, its CIDR block cannot be modified.
Description	Enter the description of the VSwitch.

5. Click OK.

17.4.2 View VSwitches

On the VSwitches tab page, you can view the information about created VSwitches.

Procedure

1. *Log on to the VPC console*.
2. On the VPC tab page, locate the relevant VPC, and click the VPC ID.
3. Click the VSwitches tab to view VSwitch information.

17.4.3 Edit VSwitch information

After you create a VSwitch, you can edit the name and description of this VSwitch.

Procedure

1. *Log on to the VPC console*.

2. Locate the relevant VPC, and click the VPC ID.
3. Click the VSwitch tab.
4. Locate the relevant VSwitch, click the icon in the Actions column, and choose Edit from the shortcut menu.
5. In the dialog box that appears, edit the name and description of the VSwitch, and click OK.

17.4.4 Delete a VSwitch

You can delete a created VSwitch.

Prerequisites

Before deleting a VSwitch, you must release or move cloud services from the VSwitch.

Procedure

1. *Log on to the VPC console*.
2. On the VPC tab page, locate the relevant VPC, and click the VPC ID.
3. Click the VSwitch tab.
4. Locate the relevant VSwitch, click the icon in the Actions column, and choose Delete from the shortcut menu.
5. In the message that appears, click OK.

17.5 VRouter and routing table

17.5.1 Overview

A VRouter is the network hub in a VPC. It is an important component of the VPC. It connects VSwitches in the VPC and serves as the gateway that connects the VPC to gateways in other networks.

A VRouter is automatically created for a VPC after the VPC is created. When the VPC is deleted, the VRouter is also deleted. A VRouter cannot be created or deleted manually. Each VRouter maintains a routing table. A VRouter forwards network traffic based on the routing entries in the routing table.

A routing table is a list of routing entries stored in the VRouter. A routing table is automatically created for a VPC after the VPC is created. When the VPC is deleted

, the routing table is also deleted. A routing table cannot be created or deleted manually.

Each item in a routing table is a routing entry. The routing entry defines the next-hop IP address for the network traffic to be routed to the specified destination CIDR block. Two types of routes are available: system routes and custom routes.

- System routes

  A system routing entry is automatically created for a VPC after the VPC is created. This routing entry defines the routes for the cloud service instances in the VPC to communicate with each other. A system routing entry is also automatically created for a VSwitch after the VSwitch is created. The CIDR block of this VSwitch is the destination. For more information, see *View VRouters and routing tables*.

- Custom routes

  You can add a custom route to forward the destination traffic to a specified next hop. For more information, see *Create a routing entry*.

17.5.2 View VRouters and routing tables

You can view created VRouters and their routing tables.

Procedure

1. *Log on to the VPC console*.
2. On the VPC tab page, locate the relevant VPC, and click the VPC ID.
3. Click the VRouter tab to view the VRouter and routing table.

17.5.3 Create a routing entry

Each item in a routing table is a routing entry. A routing entry specifies the next-hop IP address for traffic destined for a specified CIDR block. You can customize a maximum of 48 routing entries in a routing table.

Procedure

1. *Log on to the VPC console*.
2. On the VPCs tab, click a VPC ID.
3. Click the VRouter tab.
4. In the Route Table section, click Create.

5. **In the dialog box that appears, set the following parameters to create a routing entry.**

Table 17-6: Parameters for creating a routing entry

Parameter	Description
VPC CIDR Block	Enter the destination CIDR block of the routing entry. When you set this parameter, note the following limits: · The destination CIDR block of the routing entry cannot be the same as or a subset of the CIDR blocks of any VSwitches in the VPC. · The destination CIDR block of the routing entry cannot be 100.64.0.0/10 or a subset of it. · Routing entries in the same routing table cannot share the same destination CIDR block. · If the specified destination CIDR block is an IP address, the subnet mask /32 is used by default.
Next Hop Type	Select the next-hop type of the routing entry. Valid values: · ECS Instances: indicates that the requests from the destination CIDR block are forwarded to an ECS instance. · Router Interfaces: indicates that the requests from the destination CIDR block are forwarded to a specified VRouter interface. The requests are routed to the peer VRouter interface through the specified local VRouter interface. · High-Availability Virtual IPs: indicates that the requests from the destination CIDR block are forwarded to a high-availability virtual IP address.
Next Hop Type is set to ECS Instances.	
Next Hop Instance ID	Select the ECS instance that receives the forwarded traffic. When you set this parameter, note the following limits: · The next-hop ECS instance specified in the routing entry must belong to the same VPC as the routing table. · Multiple routing entries can be destined for the same ECS instance.
Next Hop Type is set to Router Interfaces.	
Route Type	Select the route type. Valid values: · Regular Route · Load Balancing Route

Parameter	Description
Router Interfaces	Select the router interface that receives the forwarded traffic.
Next Hop Type is set to High-Availability Virtual IPs.	
High-Availability Virtual IPs	Select the high-availability virtual IP address that receives the forwarded traffic.

6. Click OK.

17.6 NAT Gateway

17.6.1 Overview

NAT Gateway is an enterprise gateway for communication between VPCs and the Internet. It provides NAT proxy services (SNAT and DNAT), up to 10 Gbit/s forwarding capacity, and cross-zone disaster tolerance. You can use NAT Gateway together with a shared bandwidth package to build an enterprise gateway with high performance and flexible configuration.

NAT Gateway provides the following functions:

- Bandwidth package: You can configure a public IP address for NAT Gateway by adding a bandwidth package. A bandwidth package consists of an Internet bandwidth and a group of public IP addresses.
- EIP: You can configure a public IP address for NAT Gateway by associating an EIP with NAT Gateway.
- DNAT (port forwarding): DNAT refers to Destination Network Address Translation. NAT Gateway can DNAT inbound traffic from the Internet to ECS instances in the VPC. Both port mapping and IP mapping are supported.
- SNAT: SNAT refers to Source Network Address Translation. NAT Gateway can SNAT outbound traffic from ECS instances in the VPC to the Internet.

NAT Gateway has the following features:

- Bandwidth sharing with multiple IP addresses: All IP addresses in a bandwidth package share an Internet bandwidth.
- High performance: NAT Gateway provides a forwarding capacity of up to 10 Gbit/s per instance.

- **High availability: Based on SDN technology, NAT Gateway uses a distributed architecture that allows you to deploy multiple instances across different zones. Each instance can take over services upon failure in a certain zone.**
- **Configuration change at any time: You can change the instance specifications, bandwidth, and number of public IP addresses of NAT Gateway at any time. The change takes effect immediately.**

17.6.2 Create a NAT Gateway instance

NAT Gateway is an enterprise gateway for communication between VPCs and the Internet. You can create a NAT Gateway instance to provide NAT proxy services for ECS instances in your VPC.

Prerequisites

Before creating a NAT Gateway instance, make sure that you have created a VPC.

Procedure

1. *Log on to the VPC console*.
2. **Click the NAT Gateways tab, and then click Create Instance.**
3. **In the Create NAT Gateway dialog box that appears, set parameters for creating a NAT Gateway instance. The following table describes the parameters.**

Table 17-7: Parameters for creating a NAT Gateway instance

Parameter	Description
Region	Display the region of the NAT Gateway instance.
Department	Select the department of the NAT Gateway instance.
VPC	Select the VPC that uses the NAT Gateway instance.

Parameter	Description
Specifications	Select the specification of the NAT Gateway instance. The maximum number of SNAT connections and number of new SNAT connections per second vary depending on the specification of the NAT Gateway instance. The SNAT data throughput is not restricted by the specification of the NAT Gateway instance. · Small: The maximum number of SNAT connections is 10,000, and the number of new SNAT connections per second is 1,000. · Medium: The maximum number of SNAT connections is 50,000, and the number of new SNAT connections per second is 5,000. · Large: The maximum number of SNAT connections is 200,000, and the number of new SNAT connections per second is 10,000. Note: The maximum number of DNAT connections and DNAT data throughput are not restricted by the specification of the NAT Gateway instance.
Public IP Addresses	Select a bandwidth package or an existing EIP. · If you select a bandwidth package, you need to enter the number of public IP addresses and bandwidth. · If you select an existing EIP, you need to associate it with the NAT Gateway instance.
Instance Name	Enter the name of the NAT Gateway instance.

4. Click Create.

Result

When a NAT Gateway instance is created, the system automatically creates a port forwarding table and an SNAT table. In addition, the system automatically creates a bandwidth package based on the number of public IP addresses and bandwidth you have specified. To view the created bandwidth package, click the ID of the created

NAT Gateway instance, go to the NAT Gateways page, and then click the Bandwidth Packages tab. A maximum of four bandwidth packages can be created for a NAT Gateway instance. For more information, see *Create a bandwidth package*.

17.6.3 View a NAT Gateway instance

You can view details of a created NAT Gateway instance.

Procedure

1. *Log on to the VPC console*.
2. Click the NAT Gateways tab.
3. Click the ID of the target NAT Gateway instance. Alternatively, click , and then click Details.

17.6.4 Modify the name and description of a NAT Gateway instance

You can modify the name and description of a NAT Gateway instance.

Procedure

1. *Log on to the VPC console*.
2. Click the NAT Gateways tab.
3. Click next to the target NAT Gateway instance, and then click Change.
4. In the dialog box that appears, modify the name and description of the NAT Gateway instance, and then click OK.

17.6.5 Modify the type of a NAT Gateway instance

NAT Gateway instances can be of three types: small, medium, and large. The maximum number of SNAT connections and number of new SNAT connections per second vary depending on the type of NAT Gateway instances. The SNAT data throughput is not restricted by the type of NAT Gateway instances. The maximum number of DNAT connections and DNAT data throughput are not restricted by the type of NAT Gateway instances.

Context

The bandwidth and number of IP addresses in a shared service plan are not restricted by the type of NAT Gateway instances. A service plan with the maximum bandwidth can be configured for a small NAT Gateway instance.

Procedure

1. *Log on to the VPC console*.
2. Click the NAT Gateways tab.
3. Click the icon in the Actions column corresponding to a NAT Gateway instance, and choose Change Specification from the shortcut menu.
4. In the dialog box that appears, select a new type for the NAT Gateway instance.

   Table 17-8: Types available for a NAT Gateway instance

Type	Maximum number of SNAT connections	Number of new SNAT connections per second	Maximum number of DNAT connections	Number of new DNAT connections per second
Small	10,000	1,000	Unlimited	Unlimited
Medium	50,000	5,000	Unlimited	Unlimited
Large	200,000	10,000	Unlimited	Unlimited

5. Click OK.

17.6.6 Delete a NAT Gateway instance

You can delete a created NAT Gateway instance.

Prerequisites

Before deleting the NAT Gateway instance, you must delete the bandwidth packages of the NAT Gateway instance.

Procedure

1. *Log on to the VPC console*.
2. Click the NAT Gateways tab.
3. Click next to the target NAT Gateway instance, and then click Delete.
4. In the message that appears, click OK.

17.6.7 Bandwidth package

17.6.7.1 Create a bandwidth package

A bandwidth package encapsulates the public IP addresses and Internet bandwidth specified for a NAT Gateway instance. A bandwidth package consists of an Internet

bandwidth and a group of public IP addresses that share the bandwidth. When a NAT Gateway instance is created, the system automatically creates a bandwidth package based on the number of public IP addresses and bandwidth you have specified. After a bandwidth package is created, you can change the number of public IP addresses and bandwidth in the bandwidth package. You can also create more bandwidth packages.

Prerequisites

You have created a NAT Gateway instance. A maximum of four bandwidth packages can be created for a NAT Gateway instance.

> **Note:**
> The system automatically creates a bandwidth package based on the number of public IP addresses and bandwidth specified when you create the NAT Gateway instance.

Procedure

1. *Log on to the VPC console*.
2. Click the NAT Gateways tab, and then click the ID of the target NAT Gateway instance.
3. Click the Bandwidth Packages tab, and then click Create Bandwidth Package.
4. In the Create Bandwidth Package dialog box that appears, specify the name, zone, number of public IP addresses, and bandwidth of the bandwidth package, and then click OK.

   All IP addresses in the bandwidth package share the bandwidth. You can change the bandwidth value at any time.

17.6.7.2 View a bandwidth package

You can view information about a created bandwidth package.

Procedure

1. *Log on to the VPC console*.
2. Click the NAT Gateways tab, and then click the ID of the target NAT Gateway instance.
3. Click the Bandwidth Packages tab to view information about a created bandwidth package.

17.6.7.3 Modify the name and description of a bandwidth package

You can modify the name and description of a created bandwidth package.

Procedure

1. *Log on to the VPC console*.
2. **Click the NAT Gateways tab, and then click the ID of the target NAT Gateway instance.**
3. **Click the Bandwidth Packages tab.**
4. **In the Basic Information area, click Change.**
5. **In the dialog box that appears, modify the name and description of the bandwidth package, and then click OK.**

17.6.7.4 Modify the bandwidth of a bandwidth package

You can modify the bandwidth of a created bandwidth package.

Procedure

1. *Log on to the VPC console*.
2. **Click the NAT Gateways tab, and then click the ID of the target NAT Gateway instance.**
3. **Click the Bandwidth Packages tab.**
4. **In the Basic Information area, click** `Change Bandwidth`.
5. **In the dialog box that appears, modify the bandwidth and click OK.**

17.6.7.5 Add a public IP address to a bandwidth package

You can add a public IP address to a created bandwidth package.

Procedure

1. *Log on to the VPC console*.
2. **Click the NAT Gateways tab, and then click the ID of the target NAT Gateway instance.**
3. **Click the Bandwidth Packages tab.**
4. **In the Public IP Addresses area, click Add Public IP Addresses.**
5. **Enter the number of public IP addresses, and then click OK.**

17.6.7.6 Remove a public IP address from a service plan

You can remove a public IP address from an existing service plan.

Procedure

1. *Log on to the VPC console*.
2. **Click the NAT Gateways tab, and then click the ID of the target NAT Gateway instance.**
3. **Click the Bandwidth Packages tab.**
4. **In the Public IP Addresses section, click the icon in the Actions column corresponding to the public IP address to be removed, and choose Remove from the shortcut menu.**
5. **In the message that appears, click OK to remove the public IP address.**

17.6.7.7 Delete a service plan

You can delete a service plan that is no longer needed.

Prerequisites

Before you delete a service plan, make sure that the public IP address of the service plan is not used in any port forwarding rules or SNAT rules.

Procedure

1. *Log on to the VPC console*.
2. **Click the NAT Gateways tab, and then click the ID of the target NAT Gateway instance.**
3. **Click the Bandwidth Packages tab.**
4. **Click the Delete icon corresponding to the service plan to be deleted.**
5. **In the message that appears, click OK.**

17.6.8 DNAT table

17.6.8.1 Create a DNAT entry

NAT Gateway supports DNAT, which maps a public IP address to an ECS instance in a VPC so that the ECS instance can provide Internet services. The system automatically creates a DNAT table for each NAT Gateway instance when it is created. You can create a DNAT entry in the DNAT table to map a public IP address to an ECS instance.

Prerequisites

Create a NAT Gateway instance. When the NAT Gateway instance is created, the system automatically creates a DNAT table for it. You can create a DNAT entry so that an ECS instance in a VPC can provide Internet services.

Procedure

1. *Log on to the VPC console*.
2. Click the NAT Gateways tab, and then click the ID of the target NAT Gateway instance.
3. Click the DNAT Table tab.
4. In the DNAT Entries area, click Create DNAT Entry.
5. In the Create DNAT Entry dialog box that appears, set parameters for creating a DNAT entry. The following table describes the parameters.

Table 17-9: Parameters for creating a DNAT entry

Parameter	Description
Public IP Address	Select a public IP address type. · EIP: Click EIP, and then select an EIP. · Bandwidth Packages: Click Bandwidth Packages, and then select a bandwidth package and a public IP address. Note: The same public IP address cannot be used for both SNAT and DNAT.
Private IP Address	Specify the private IP address of the ECS instance to which the public IP address maps. You can specify a private IP address in either of the following ways: · Custom: Click Custom, and then enter the private IP address of the ECS instance to be mapped. · Select from ECS IP Addresses: Click Select from ECS IP Addresses, and select the ECS instance to be mapped from the drop-down list. The system automatically fills in the private IP address of the selected ECS instance. Note: The ECS instance must belong to the same VPC as the NAT Gateway instance.

Parameter	Description
Port Settings	Select a port setting method. Valid values: · All: This method is called IP mapping. The selected ECS instance exclusively occupies the specified public IP address to communicate with the Internet in both the inbound and outbound directions. This is equivalent to associating an EIP with the ECS instance. · Specific Port: This method is called port mapping. You must specify the port to be mapped. The NAT Gateway instance forwards requests in compliance with the specified protocol and from the specified port to the specified port of the target ECS instance through the specified public IP address.
The following parameters are available only when you set the port setting method to Specific Port.	
Public Network Port	Enter the public port for receiving data forwarded by the NAT Gateway instance.
VPC Network Port	Enter the private port for receiving data.
Protocol Type	Select the protocol type of the received data.

6. Click OK.

17.6.8.2 View the DNAT table

You can view details of the DNAT table of a NAT Gateway instance.

Procedure

1. *Log on to the VPC console*.
2. Click the NAT Gateways tab, and then click the ID of the target NAT Gateway instance.
3. Click the DNAT Table tab to view the DNAT table.

17.6.8.3 Modify a DNAT entry

You can modify a DNAT entry on the DNAT Table tab.

Procedure

1. *Log on to the VPC console*.

2. Click the NAT Gateways tab, and then click the ID of the target NAT Gateway instance.
3. Click the DNAT Table tab.
4. Click next to the target DNAT entry in the Actions column, and then click Change to modify the DNAT entry.

   For more information, see *Create a DNAT entry*.

17.6.8.4 Delete a DNAT entry

You can delete a DNAT entry on the DNAT Table tab.

Procedure

1. *Log on to the VPC console*.
2. Click the NAT Gateways tab, and then click the ID of the target NAT Gateway instance.
3. Click the DNAT Table tab.
4. Click next to the target DNAT entry in the Actions column, and then click Delete.
5. In the message that appears, click OK.

17.6.9 SNAT table

17.6.9.1 Create an SNAT entry

NAT Gateway supports SNAT, which allows an ECS instance without a public IP address in a VPC to access the Internet.

Prerequisites

Create a NAT Gateway instance. When the NAT Gateway instance is created, the system automatically creates an SNAT table for it. You can create an SNAT entry so that an ECS instance in a VPC can access the Internet. For more information, see *Create an SNAT entry*.

Procedure

1. *Log on to the VPC console*.
2. Click the NAT Gateways tab, and then click the ID of the target NAT Gateway instance.
3. Click the SNAT Table tab.

4. In the SNAT Entries area, click Create SNAT Entry.
5. In the Create SNAT Entry dialog box that appears, set parameters for creating an SNAT entry. The following table describes the parameters.

   Table 17-10: Parameters for creating an SNAT entry

Parameter	Description
Switch	Select the VSwitch of the ECS instance to be granted access to the Internet. **Note:** When an ECS instance associated with a public IP address (for example, an EIP) in a VSwitch accesses the Internet, the ECS instance preferentially uses the public IP address without activating the SNAT function of NAT Gateway.
Switch Network Segment	Display the CIDR block of the selected VSwitch.
Public IP Address	Select a public IP address type. · EIP: Click EIP, and then select an EIP. · Bandwidth Packages: Click Bandwidth Packages, and select a bandwidth package and a public IP address. **Note:** The same public IP address cannot be used for both SNAT and DNAT.

6. Click OK.

17.6.9.2 View the SNAT table

You can view information about the SNAT table on the SNAT Table tab.

Procedure

1. *Log on to the VPC console*.
2. Click the NAT Gateways tab, and then click the ID of the target NAT Gateway instance.
3. Click the SNAT Table tab to view the SNAT table.

17.6.9.3 Modify an SNAT entry

You can modify an SNAT entry on the SNAT Table tab.

Procedure

1. *Log on to the VPC console*.
2. **Click the NAT Gateways tab, and then click the ID of the target NAT Gateway instance.**
3. **Click the SNAT Table tab.**
4. **Click the icon in the Actions column corresponding to the SNAT entry to be modified, and choose Update SNAT Rule from the shortcut menu.**
5. **In the dialog box that appears, modify the SNAT entry and click OK.**

17.6.9.4 Delete an SNAT entry

You can delete an SNAT entry on the SNAT Table tab.

Procedure

1. *Log on to the VPC console*.
2. **Click the NAT Gateways tab, and then click the ID of the target NAT Gateway instance.**
3. **Click the SNAT Table tab.**
4. **Click the icon in the Actions column corresponding to the SNAT entry to be deleted, and choose Delete from the shortcut menu.**
5. **In the message that appears, click OK.**

17.6.10 EIP

17.6.10.1 Associate an EIP with a NAT Gateway instance

You can configure a public IP address for a NAT Gateway instance by associating an EIP with it.

Prerequisites

You have created a NAT Gateway instance and an EIP. You can associate a maximum of five EIPs with a NAT Gateway instance.

Procedure

1. *Log on to the VPC console*.

2. Click the NAT Gateways tab, and then click the ID of the target NAT Gateway instance.
3. Click the EIP tab, and then click Bind to EIP.
4. In the Bind to EIP dialog box that appears, select the project and line type, and then click Search.
5. Select the EIP to be associated, and then click OK.

17.6.10.2 Disassociate an EIP from a NAT Gateway instance

You can disassociate an EIP from a NAT Gateway instance.

Prerequisites

Make sure that the EIP to be disassociated is not occupied by any SNAT or DNAT entries.

Procedure

1. *Log on to the VPC console*.
2. Click the NAT Gateways tab, and then click the ID of the target NAT Gateway instance.
3. Click the EIP tab.
4. Click next to the target EIP in the Actions column, and then click Unbind.
5. In the message that appears, click OK.

17.7 VRouter interface

17.7.1 Overview

A VRouter interface is a virtual device that can build a communication channel and control the channel working status. You can use VRouter interfaces to connect two VPCs for private network communication.

Alibaba Cloud abstracts the process of building a private network communication channel between two VPCs as follows: Create interfaces on the VRouters in both VPCs, and connect the VRouter interfaces to build a communication channel. The two VRouters can forward messages to each other through this channel. In this way, resources such as ECS instances deployed in the two VPCs can communicate with each other over the private network.

When two VRouter interfaces are connected, one acts as the connection initiator , and the other acts as the connection acceptor. The concepts of the initiator and acceptor are distinguished only in the process of establishing a connection. During network communication, the communication link is bidirectional, and each VRouter interface functions as both the initiator and acceptor. The following table compares the initiator and acceptor.

Table 17-11: Comparison between the initiator and acceptor

Item	Initiator	Acceptor
Whether the peer information needs to be configured on the local VRouter interface before a connection is established	Yes	Yes
Whether the local VRouter interface can initiate a connection	Yes	No
Whether the local VRouter interface can send messages to the peer after the connection is established	Yes	Yes
Whether the role of the local VRouter interface can be modified after the VRouter interface is created	No	No

When a VRouter interface is created, it may be of any of the following statuses (excluding the intermediate statuses such as Connecting and Activating):

- **Idle: indicates that the local and peer VRouter interfaces have not been connected. In this case, the VRouter interface with the initiator role needs to initiate a connection.**
- **Active: indicates that the VRouter interface is connected. If routing information is correctly configured, the VRouter interface can send data properly.**
- **Inactive: The activated VRouter interface is deactivated. In this case, no data passes through the VRouter interface.**

17.7.2 Create a VRouter interface

You can create an interface on the VRouters of two VPCs respectively so that the cloud resources in the two VPCs can communicate with each other. When creating a VRouter interface, you must specify the role of the VRouter interface as the

initiator or acceptor. The VRouter interface with the initiator role initiates a connection.

Procedure

1. *Log on to the VPC console*.
2. Click the Router Interfaces tab, and then click Create.
3. In the Create Router Interface dialog box that appears, set parameters for creating a VRouter interface. The following table describes the parameters.

Table 17-12: Parameters for creating a VRouter interface

Parameter	Description
Basic Settings	
Role	Select the connection role of the VRouter interface. The VRouter interface with the initiator role can initiate a connection. **Note:** You cannot modify the role of a VRouter interface after the VRouter interface is created.
Initiator Configuration	
Router Type	Select the type of the VRouter.
Region	Select the region of the local VPC.
Department	Select the department of the local VPC.
VPC	Select a VPC as the local end in the VPC connection.
VRouter	Display the VRouter ID of the local VPC.
Specifications	Select the specification of the VRouter interface.
Acceptor Configuration	
Router Type	Display the type of the VRouter.

Parameter	Description
Region	Select the region of the peer VPC to be connected.
Department	Select the department of the peer VPC.
VPC	Select the peer VPC to be connected.
VRouter	Display the VRouter ID of the peer VPC.

4. Click Create.

17.7.3 Modify information of the local VRouter interface

You can modify the name and description of the local VRouter interface on the Router Interfaces tab.

Procedure

1. *Log on to the VPC console*.
2. Click the Router Interfaces tab.
3. Click next to the target VRouter interface, and then click Modify Local Interface.
4. In the dialog box that appears, modify the name and description of the local VRouter interface, and then click Change.

17.7.4 Modify information of the peer VRouter interface

You can modify the name and description of the peer VRouter interface on the Router Interfaces tab.

Procedure

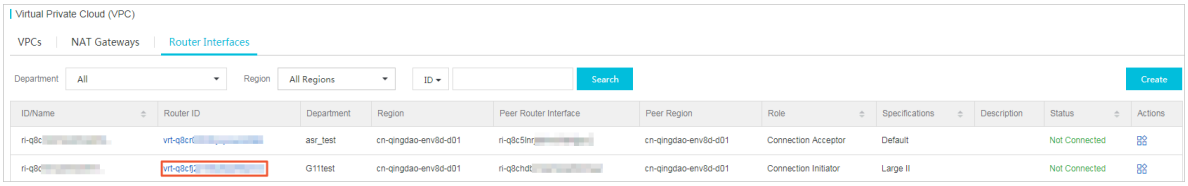
1. *Log on to the VPC console*.
2. Click the Router Interfaces tab.
3. Click next to the target VRouter interface, and then click Modify Peer Interface.
4. In the dialog box that appears, select the department and VPC of the peer VRouter interface, enter the VRouter interface ID, and then click Change.

17.7.5 Create a route

You can create a route for a VRouter interface on the Router Interfaces tab.

Procedure

1. *Log on to the VPC console*.
2. **Click the Router Interfaces tab.**
3. **Click the VRouter ID of the target VRouter interface.**



4. **Click the VRouter tab.**
5. **In the Route Table area, click Create.**
6. **In the dialog box that appears, set parameters for creating a route. The following table describes the parameters.**

   Table 17-13: Parameters for creating a route

Parameter	Description
VPC CIDR Block	Enter the CIDR block of the peer VPC.
Next Hop Type	Select Router Interfaces.
Router Interfaces	Select the VRouter interface of the local VPC.

7. **Click OK.**

17.7.6 Initiate a connection

After being activated, the VRouter interface with the initiator role can initiate a connection.

Procedure

1. *Log on to the VPC console*.
2. **Click the Router Interfaces tab.**
3. **Click next to the target VRouter interface, and then click Initiate Connection.**

17.7.7 Activate a VRouter interface

You can activate a VRouter interface in the Inactive status. After being activated, the VRouter interface can send data properly.

Procedure

1. *Log on to the VPC console*.
2. **Click the Router Interfaces tab.**
3. **Click next to the target VRouter interface, and then click Activate.**
4. **In the message that appears, click OK.**

17.7.8 Deactivate a VRouter interface

You can deactivate a VRouter interface in the Active status. After the VRouter interface is deactivated, no data passes through it.

Procedure

1. *Log on to the VPC console*.
2. **Click the Router Interfaces tab.**
3. **Click next to the target VRouter interface, and then click Freeze.**
4. **In the message that appears, click OK.**

17.7.9 Delete a VRouter interface

You can delete a VRouter interface in the Idle or Inactive status.

Procedure

1. *Log on to the VPC console*.
2. **Click the Router Interfaces tab.**
3. **Click next to the target VRouter interface, and then click Delete.**
4. **In the message that appears, click OK.**

17.8 EIP

17.8.1 Apply for an EIP

An EIP is a public IP address resource that you can purchase and use independently. You can dynamically associate an EIP with ECS instances in different VPCs. No downtime is required when you associate or disassociate an EIP. You can

apply for an EIP and associate it with an ECS instance in your VPC. In this way, the ECS instance can communicate with the Internet.

Context

An EIP is a NAT IP address mapped to the private network interface controller (NIC) of the associated ECS instance through NAT. Therefore, an ECS instance associated with an EIP can directly use the EIP for Internet communication. The EIP cannot be read on the NIC.

Note:
Currently, you can associate EIPs only with ECS instances.

Procedure

1. Log on to Apsara Stack console.
2. In the top navigation bar, choose > Compute, Storage & Networking > Elastic IP Address.
3. Click Apply for EIP.
4. On the page that appears, select the region, line type, department, and project of the required EIPs, set the peak bandwidth and number of EIPs, and then click Create.

17.8.2 Associate an EIP with an ECS instance

You can associate an EIP with an ECS instance in your VPC so that the ECS instance can communicate with the Internet.

Prerequisites

Note the following limits before you associate an EIP with an ECS instance:

- Only an EIP in the Available status can be associated.
- Currently, the network type of the ECS instance to be associated with an EIP must be VPC. In addition, the ECS instance and EIP must belong to the same region and department.
- An ECS instance can be associated with only one EIP. An EIP can be associated with only one ECS instance.
- You can associate an EIP only with an ECS instance in the Running or Stopped status.

- You cannot associate or disassociate an EIP that is locked for the sake of security.

Procedure

1. Log on to Apsara Stack console.
2. In the top navigation bar, choose  > Compute, Storage & Networking > Elastic IP Address.
3. Click  next to the target EIP, and then click Bind.
4. In the dialog box that appears, select the type and ID of the ECS instance to be associated, and then click OK.

17.8.3 Modify the bandwidth of an EIP

You can modify the bandwidth of an EIP.

Procedure

1. Log on to Apsara Stack console.
2. In the top navigation bar, choose  > Compute, Storage & Networking > Elastic IP Address.
3. Click  next to the target EIP, and then click Change.
4. In the dialog box that appears, modify the bandwidth of the EIP and click OK.

17.8.4 Disassociate an EIP from an ECS instance

When you do not need to access the Internet, you can disassociate an EIP from an ECS instance at any time.

Prerequisites

You can disassociate an EIP only from an ECS instance in the Running or Stopped status.

Procedure

1. Log on to Apsara Stack console.
2. In the top navigation bar, choose  > Compute, Storage & Networking > Elastic IP Address.
3. Click  next to the target EIP, and then click Unbind.

4. In the dialog box that appears, click OK.

   If the EIP status changes to `Available`, the EIP is successfully disassociated.

17.8.5 Delete an EIP

You can delete EIP resources that are no longer in use.

Prerequisites

Make sure that the EIP to be deleted is not associated with any ECS instances. Otherwise, the EIP cannot be deleted.

Procedure

1. Log on to Apsara Stack console.
2. In the top navigation bar, choose  > Compute, Storage & Networking > Elastic IP Address.
3. Click  next to the target EIP, and then click Delete.

17.9 High-Availability Virtual IP

17.9.1 Overview

High-Availability Virtual IP Address (HaVip) is a private IP resource which can be independently created and released. After you bind an HaVip instance to an ECS instance, the ECS instance can use the ARP protocol to advertise the HaVip address.

Features

HaVip provides the following features:

- An HaVip instance can be bound to a maximum of two ECS instances. After you bind an HaVip instance to an ECS instance, the ECS instance can use the ARP protocol to advertise the HaVip address.
- In addition to an ordinary private IP address, the ECS instance can advertise multiple HaVip addresses and therefore obtain more private IP addresses.
- By advertising HaVip addresses, the ECS instance can implement VRRP-based HA solutions using open-source programs such as Keepalived and Heartbeat.

 **Note:**

HaVip does not support multicasting and broadcasting. Therefore, you must set the heartbeat mode in Keepalived to unicast. When Keepalived is used, Keepalived advertises HaVip addresses. You must remove the virtual IP addresses that have been manually advertised from the NIC.

- **An HaVip instance can be bound to an EIP. After the HaVip instance is bound to a new ECS instance, messages sent to the EIP can be redirected to the new ECS instance.**

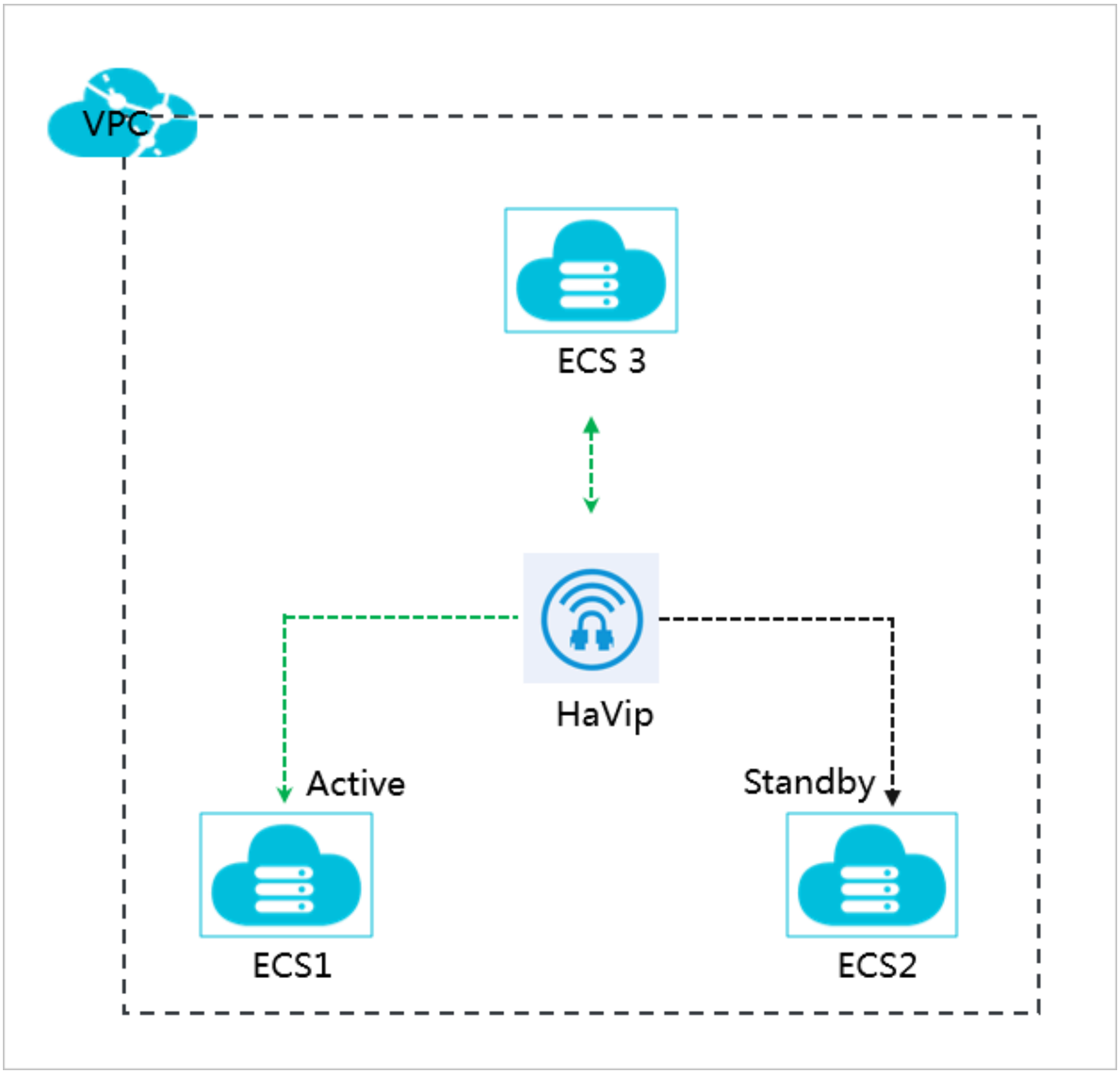
Scenarios

- **Scenario 1: Internal network-oriented HA solutions**

  HA solutions are provided using open-source programs such as Keepalived and Heartbeat.

  In the following figure, based on the HaVip instance, two ECS instances use Keepalived to implement an internal network-oriented HA solution. Other ECS instances in the same VPC can also use the HA solution. The private IP address

of the HaVip instance is used as the endpoint. When ECS 1 fails, the service is switched over to ECS 2 with the same endpoint.

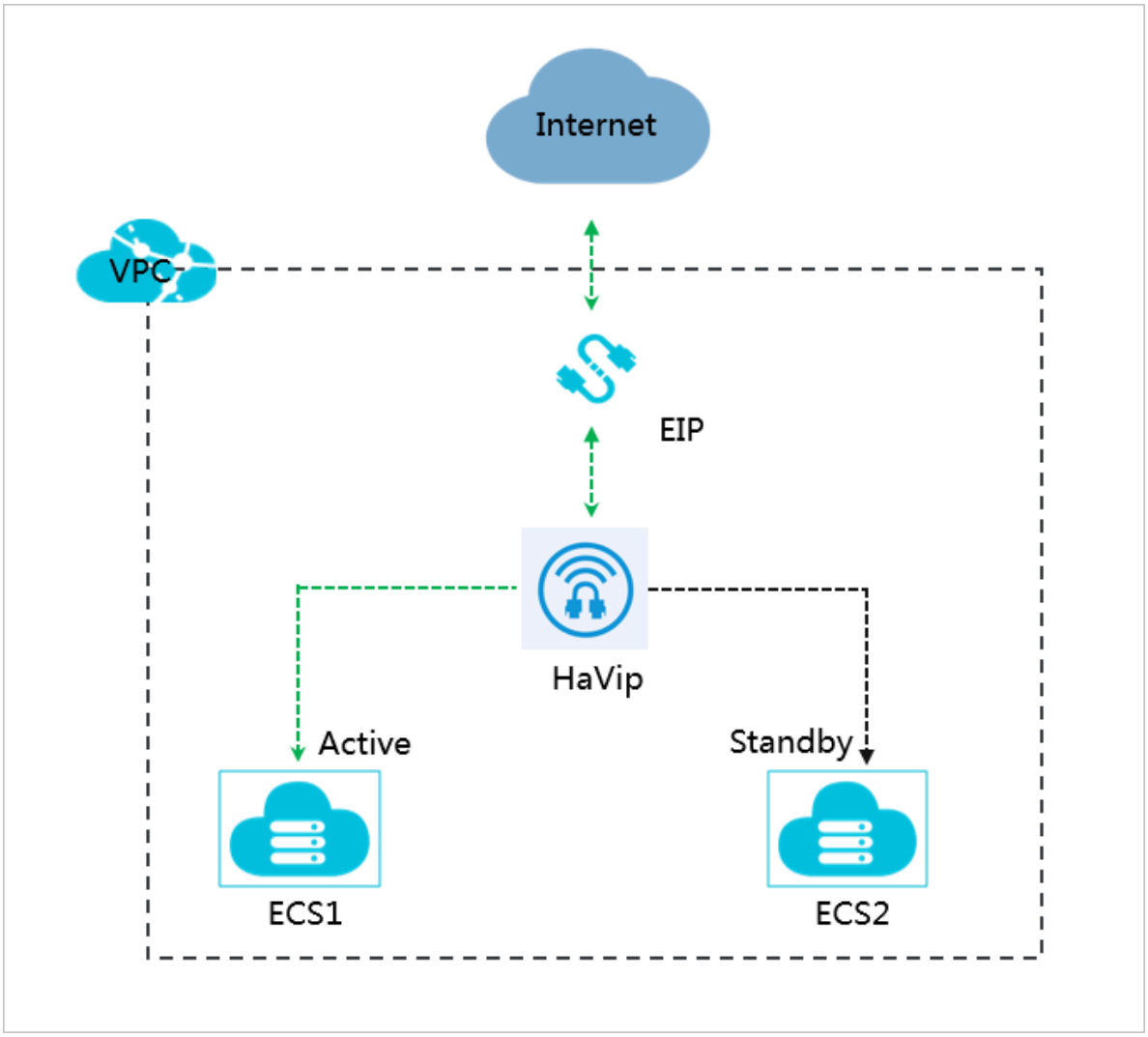


- Scenario 2: Public network-oriented HA solutions

  HA solutions are provided using open-source programs such as Keepalived and Heartbeat.

  In the following figure, based on the HaVip instance which is bound to an EIP, two ECS instances use Keepalived to implement a public network-oriented HA

solution. When ECS 1 fails, the service is switched over to ECS 2 with the same endpoint.

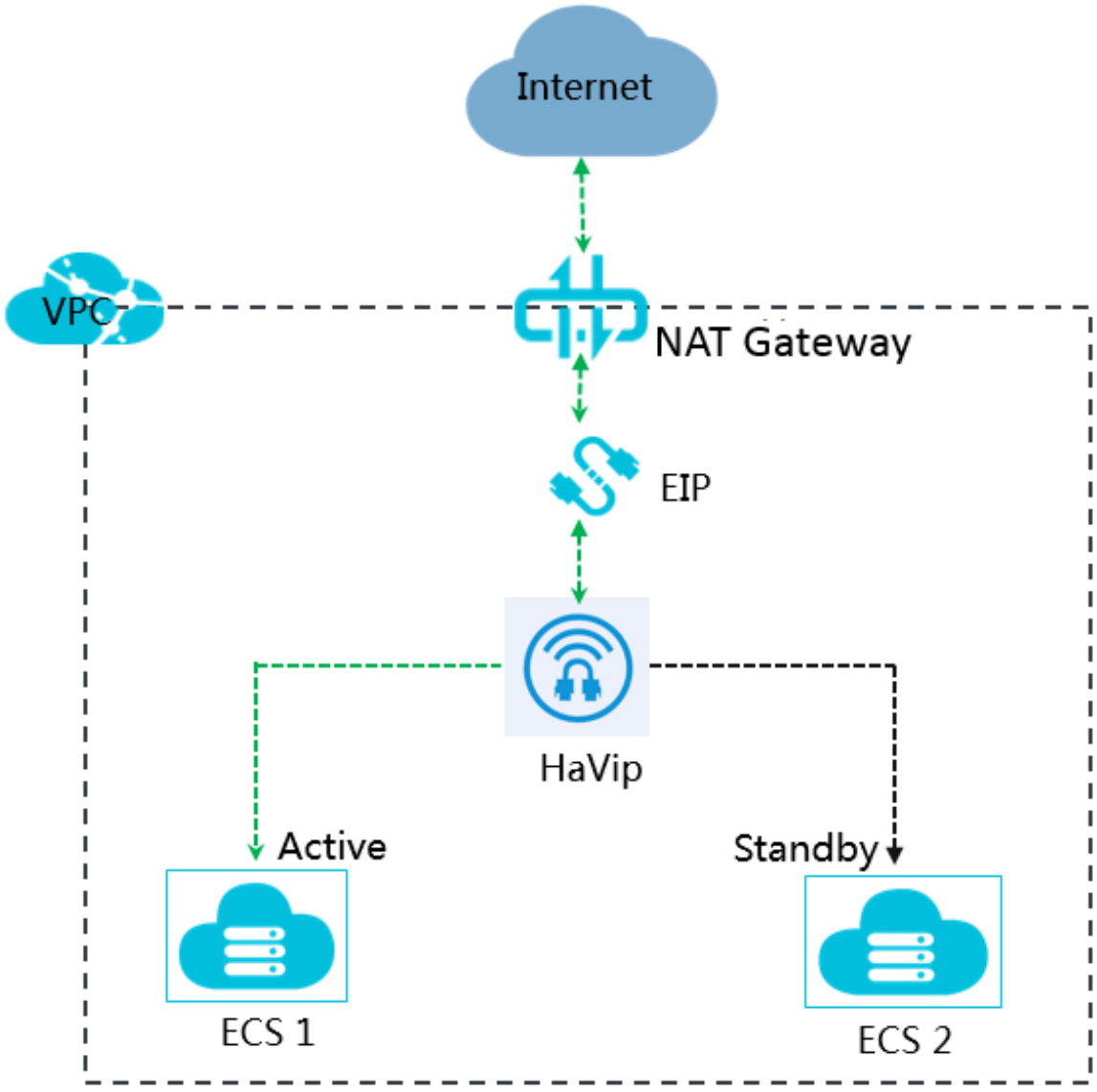


- Scenario 3: User-created NAT gateway HA solutions

  After an NAT gateway is created, other ECS instances in the same VPC can use the NAT gateway to access the public network.

  In the following figure, open-source programs such as Keepalived and Heartbeat can be used to implement a user-created SNAT gateway HA solution. The private

IP address of the HaVip instance is used as the endpoint of the SNAT gateway. When ECS 1 fails, the service is switched over to ECS 2 with the same endpoint.



17.9.2 Create an HaVip instance

High-Availability Virtual IP Address (HaVip) is a private IP resource which can be independently created and released.

Procedure

1. *Log on to the VPC console*.
2. **Click the High-Availability Virtual IPs tab and click Create Instance.**

3. **On the HAVIP Instance Creation page, configure the following parameters.**

   Table 17-14: Parameters for creating an HaVip instance

Parameter	Description
Basic Settings	
Department	**The department to which the HaVip instance belongs.**
Project	**The project to which the HaVip instance belongs.**
Region	**The region where the HaVip instance is located.**
Network Configuration	
VPC	**The VPC to which the HaVip instance belongs.**
VSwitch	**The VSwitch to which the HaVip instance belongs.**
Private IP Address	**The private IP address of the HaVip instance.** **You must select a private IP address which is still available in the CIDR block of the VSwitch. If no IP address is specified, an available private IP address in the CIDR block of the VSwitch will be automatically allocated to the HaVip instance.**

4. **Click Create.**

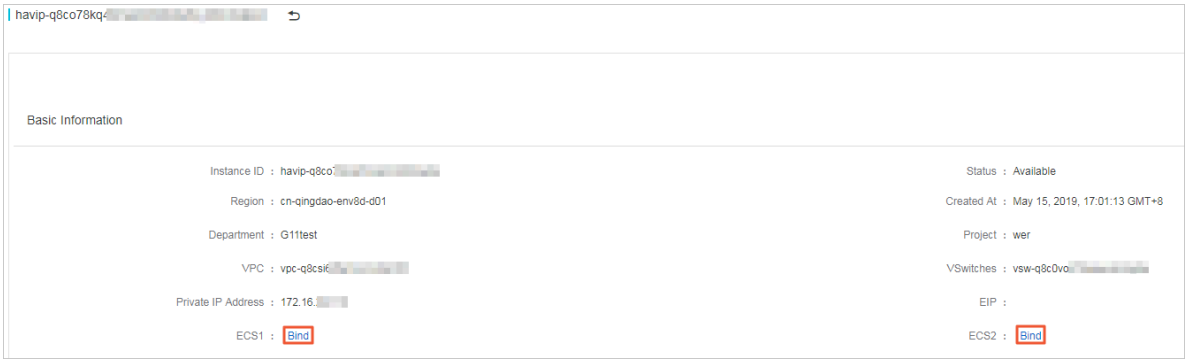
17.9.3 Bind an ECS instance

You can bind HaVip instances to an ECS instance and use open-source programs such as Keepalived and Heartbeat to implement high availability. You can bind one HaVip instance to a maximum of two ECS instances.

Procedure

1. *Log on to the VPC console*.
2. **Click the High-Availability Virtual IPs tab and click the ID of an HaVip instance.**

3. In the Basic Information section, click Bind corresponding to ECS1 and ECS2.



4. In the dialog box that appears, select the ECS instance to be bound and click OK.

17.9.4 Unbind an ECS instance

You can unbind an HaVip instance from an ECS instance. After that, the ECS instance cannot use ARP to advertise the HaVip address.

Procedure

1. *Log on to the VPC console*.
2. Click the High-Availability Virtual IPs tab and click the ID of an HaVip instance.
3. In the Basic Information section, click Unbind corresponding to ECS1 and ECS2.
4. In the message that appears, click OK.

17.9.5 Bind an EIP

You can bind HaVip instances to an EIP to provide services over the public network.

Procedure

1. *Log on to the VPC console*.
2. Click the High-Availability Virtual IPs tab.
3. Find the HaVip instance to be bound to an EIP. Click the icon in the Actions column and choose Bind to EIP from the shortcut menu.
4. Select the EIP to be bound and click OK.

17.9.6 Unbind an EIP

In the High-Availability Virtual IPs console, you can unbind EIPs that are no longer needed.

Procedure

1. *Log on to the VPC console*.

2. Click the High-Availability Virtual IPs tab.
3. Find the HaVip instance to be unbound from an EIP. Click the icon in the Actions column and choose Unbind from EIP from the shortcut menu.
4. In the message that appears, click OK.

17.9.7 Delete an HaVip instance

In the High-Availability Virtual IPs console, you can delete HaVip instances that are no longer needed.

Procedure

1. *Log on to the VPC console*.
2. Click the High-Availability Virtual IPs tab.
3. Find the HaVip instance to be deleted. Click the icon in the Actions column and choose Delete from the shortcut menu.

   Note:
   Make sure that the HaVip instance has not been bound to any cloud resources.
4. In the message that appears, click OK.

17.10 Internet access

A VPC is an isolated network environment. By default, the cloud resources in a VPC cannot access the Internet or be accessed through the Internet. Alibaba Cloud VPC provides NAT Gateway and EIP functions to enable the cloud resources in the VPC to communicate with the Internet.

- EIP: You can associate an EIP with an ECS instance in your VPC so that the ECS instance can communicate with the Internet. For more information, see *Associate an EIP with an ECS instance*.
- NAT Gateway:
  - Port forwarding table: To enable an ECS instance in your VPC to provide Internet services, you can create a port forwarding entry in the port

forwarding table to map a public IP address to the ECS instance. For more information, see *Create a DNAT entry*.

- SNAT table: To enable an ECS instance in your VPC to access the Internet, you can create an SNAT entry in the SNAT table. For more information, see *Create an SNAT entry*.

17.11 VPC connection

You can connect VPCs through VRouter interfaces.

Context

This topic uses the following two VPCs to describe how to connect VPCs through VRouter interfaces.

Table 17-15: VPC information

VPC name	CIDR block	Department
VPC_1	192.168.0.0/16	First-level department
VPC_2	172.16.0.0/12	Second-level department

First, you need to create a VRouter interface for the two VPCs respectively, and set one VRouter interface as the connection initiator and the other as the connection acceptor. Then, use the VRouter interface with the initiator role to initiate a request to connect the two VRouter interfaces. Finally, you need to configure a route for the two VRouter interfaces. For more information, see *VRouter interface*.

Procedure

1. *Log on to the VPC console*.
2. In the top navigation bar, choose  > Compute, Storage & Networking > Virtual Private Cloud.
3. Click the Router Interfaces tab, and then click Create.

4. **In the Create Router Interface dialog box that appears, set parameters for creating a VRouter interface for the two VPCs respectively.**
   - **Basic Settings**

     Role: Select Create Initiator and Acceptor.
   - **Initiator Configuration**
     - **Router Type: Select VPC.**
     - **Region: Select the region of the local VPC.**
     - **Department: Select the first-level department.**
     - **VPC: Select a VPC as the local end. In this example, select VPC_1 (192.168.0.0/16).**
     - **Specifications: Select the specification of the VRouter interface.**
   - **Acceptor Configuration**
     - **Router Type: Select VPC Router.**
     - **Region: Select the region of the peer VPC to be connected.**
     - **Department: Select the second-level department.**
     - **VPC: Select a VPC as the local end. In this example, select VPC_2 (172.16.0.0/12).**
5. **Click Create.**
6. **Create a route for the two VRouter interfaces.**
   a) **Click the VRouter ID corresponding to the VRouter interface with the initiator role in VPC_1.**
   b) **Click the VRouter tab.**
   c) **In the Route Table area, click Create and set parameters for creating a route. The following table describes the parameters.**

      Table 17-16: Parameters for creating a route for the VRouter interface with the initiator role

Parameter	Description
VPC CIDR Block	Enter the CIDR block of the peer VPC. In this example, enter 172.16.0.0/12.
Next Hop Type	Select Router Interfaces.

Parameter	Description
Router Interfaces	Select the VRouter interface of the local VPC.

d) Repeat the preceding steps to create a route for the VRouter interface with the acceptor role. The following table describes the parameters.

Table 17-17: Parameters for creating a route for the VRouter interface with the acceptor role

Parameter	Description
VPC CIDR Block	Enter the CIDR block of the peer VPC. In this example, enter 192.168.0.0/16.
Next Hop Type	Select Router Interfaces.
VRouter interface	Select the VRouter interface of the local VPC.

7. On the Router Interfaces page, click next to the VRouter interface of VPC_1 in the Actions column, and then click Initiate Connection.

   Only the VRouter interface with the initiator role can initiate a connection. If the connection is successful, the VRouter interfaces enter the Active status.

Result

When the configuration is completed, the cloud resources of VPC_1 and VPC_2 can communicate with each other.

18 Log Service

18.1 What is Log Service?

Log Service is a one-stop service designed to manage log data. You can use Log Service to perform operations on log data such as collection, query, analysis, and consumption.

Log Service has been used in various big data scenarios within Alibaba Group. You can use Log Service to collect, consume, query, and analyze log data without performing any programming. It helps increase O&M efficiency and build capabilities to process large-volume logs in the data technology (DT) era.

Log Service provides you with the following features:

- **Log collection: Log Service allows you to collect various formats of log data such as events, binary logs, and text logs in real time through multiple methods, such as Logtail and JS.**
- **Query and analysis: Log Service provides real-time query and analysis for the collected log data, and allows you to create visual charts and dashboards based on analysis results.**
- **Status alert: Log Service allows you to regularly execute, query, and analyze statements based on query and analysis features. When query results meet alert conditions, real-time alerts are reported based on pre-configured alert tasks.**
- **Real-time consumption: Log Service provides real-time consumption interfaces for log data collected to the server.**

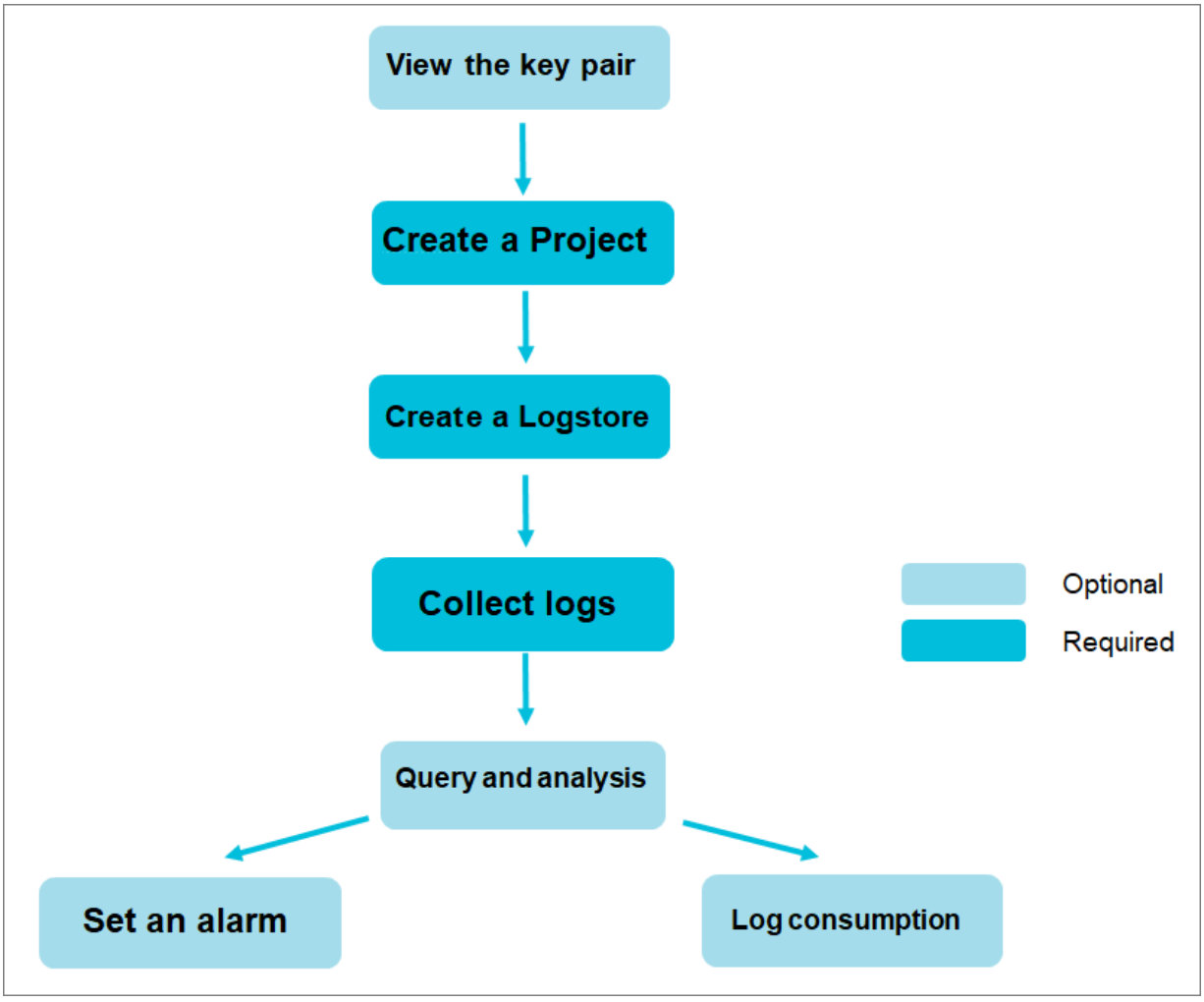
18.2 Quick start

18.2.1 Procedure

This topic describes how to use Log Service to quickly create a project and a Logstore, as well as how to collect log data.

For more information, see *Figure 18-1: Log Service flowchart*.

Figure 18-1: Log Service flowchart



1. **(Optional)** *View an AccessKey pair*.

   An AccessKey pair is essential for you to use Log Service through the API or SDKs. Make sure that you have an AccessKey pair.

2. *Create a project*.

   Create a project in the specified region and add a description.

3. *Create a Logstore*.

   Create a Logstore for the project and specify the number of shards.

4. *Collect text logs*.

   Select a method for log data collection as needed. Text log collection is used as an example.

5. *Configure an index* and query and analyze logs.

   Log Service supports *real-time query* and *analysis* for a large number of logs. With indexing enabled, you can query and analyze logs in real time.

6. *Configure an alert*.

   Log Service can trigger alerts based on the log query results. You can configure rules to send alert details by using a custom webhook method.

7. *Real-time subscription and consumption* of logs.

   Log Service enables you to consume logs by using multiple methods such as *Spark Streaming client*, *Storm spout*, and *Flink connector*.

18.2.2 Log on to the Log Service console

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://`IP address or domain name of the Apsara Stack console`/manage.
- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://`IP address or domain name of the Apsara Stack console`/manage, and then press Enter.

3. Enter the correct username and password.
   - The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click LOGIN to go to the Dashboard page.
5. Click  , and choose Compute, Storage & Networking > Log Service.
6. On the Log Service homepage, click Go to Console in the upper-right corner.
7. Set Region and Department, and then click SLS to log on to the Log Service console.

18.2.3 View an AccessKey pair

You can obtain an AccessKey pair in the Apsara Stack console.

Obtain an AccessKey pair of a department

Follow these steps:

1. Log on to the Apsara Stack console as an administrator.
2. In the top navigation bar, click the  icon, and then choose User Center > Department Management.
3. Select a department and click Create AccessKey to obtain an AccessKey pair of this department.

> **Note:**
> The AccessKey pair is automatically allocated to a level-1 department. The sub-departments use the same AccessKey pair as their level-1 department.

Obtain an AccessKey pair of a personal account

Follow these steps:

1. **Log on to the Apsara Stack console as an administrator.**
2. **In the upper-right corner of the Web page, click your avatar and select Personal Information.**

   The system displays your AccessKey ID.
3. **Click Show under AccessKey Secret to view the AccessKey Secret.**

18.2.4 Create a project

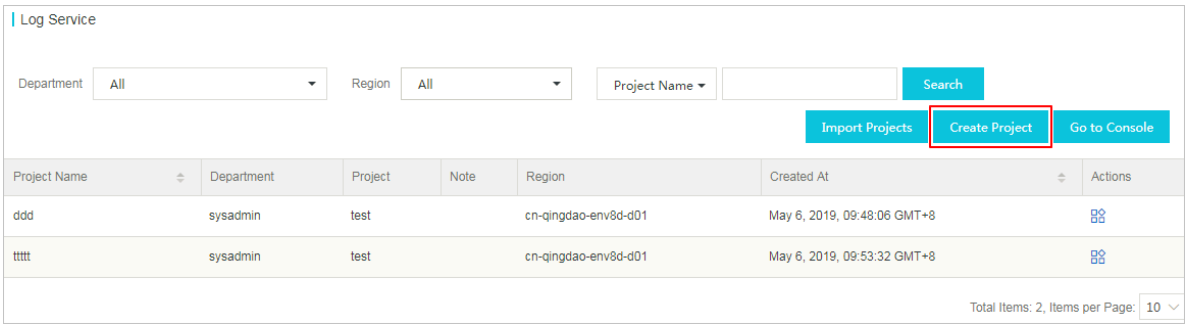
You can create a project on the Log Service homepage.

Context

Up to 10 projects can be created for each department.

Procedure

1. *Log on to the Apsara Stack console and go to the Log Service homepage*.
2. **Click Create Project in the upper-right corner.**

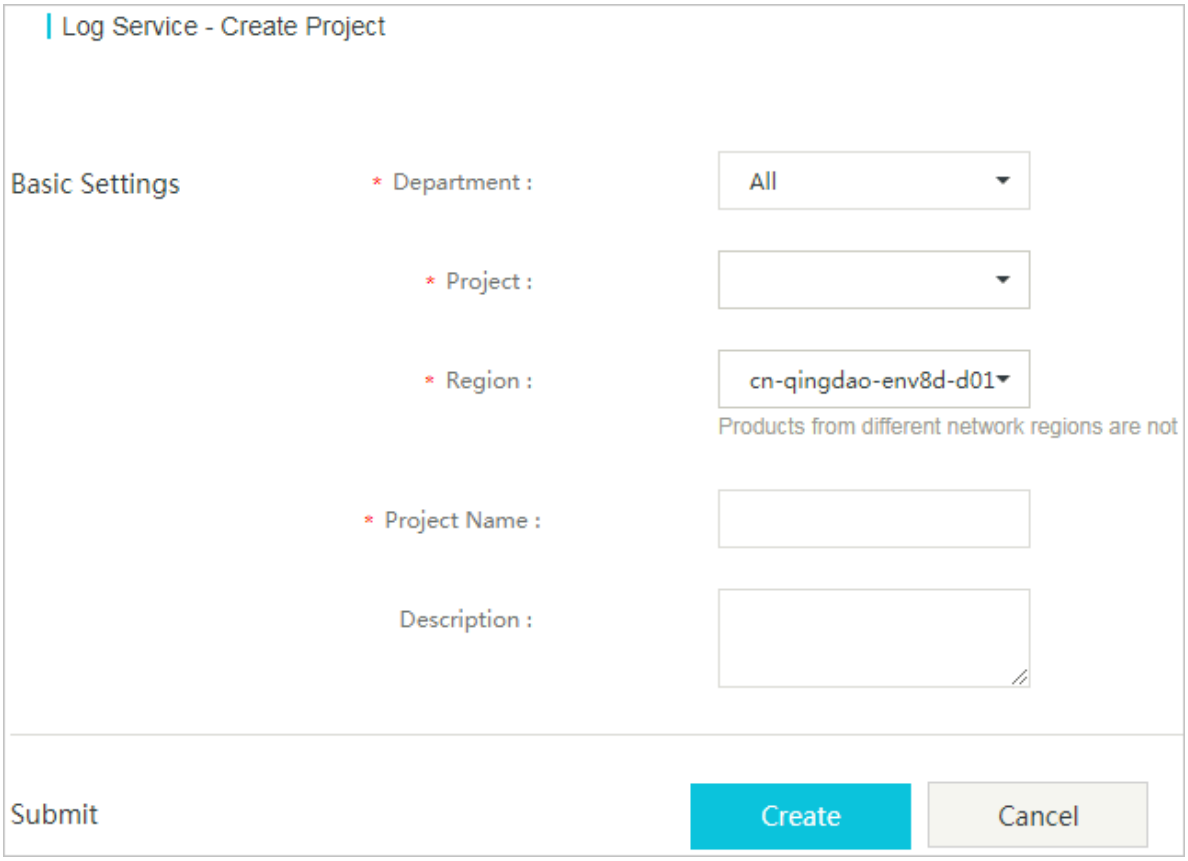


3. **Configure Basic Settings.**

Parameter	Description
Department	**The department to which the project belongs.**
Project	**The project to which the project to be created belongs.**

Parameter	Description
Region	The region of the project. When you create a project, you must select a region for the project based on log sources and other related conditions. The products and services in different regions cannot communicate with each other through the internal network. To collect logs from an ECS instance, you must create a project in the same region as the ECS instance. After a project is created, its region cannot be changed. A project cannot be migrated between different regions. Proceed with caution when selecting the region.
Project Name	The name of the project. It must be 3 to 63 characters in length and can contain only lowercase letters, digits, and hyphens (-). It must start and end with a lowercase letter or digit.  Note: · After a project is created, its name cannot be changed. · The project name must be globally unique. If the project name that you entered already exists, the Project already exists message appears. You can enter another project name and try again.

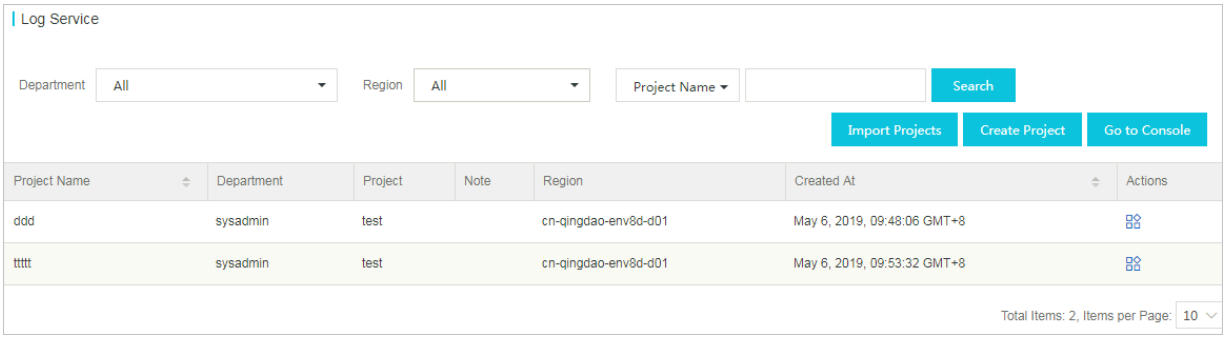
Parameter	Description
Description	The description of the project. After the project is created, the description is displayed on the Projects page. To modify the description, go to the Projects page, find the target project, click the management icon in the Actions column corresponding to the project, and then choose Change Description from the shortcut menu.



4. Click Create.

Result

You can view the created project on the Log Service homepage or in the Log Service console.



You can also set Department, Region, and Project Name, and click Search to quickly find the specified project in the project list.

18.2.5 Create a Logstore

You can use the Log Service console or API to create a Logstore. For more information about how to create a Logstore by using the Log Service API, see the following topic in *Log Service API Reference:* CreateLogstore.

Context

A Logstore is a set of resources created in a project. All the data in a Logstore comes from the same source. The collected log data is queried, analyzed, and delivered by Logstore.

- If your project is created in the Log Service console, it is automatically created in a level-1 department and does not belong to any project. If you want to import your project to the Apsara Stack console, you must migrate your project to another project by changing its ownership.
- Up to 100 Logstores can be created for each project in Log Service.

Procedure

1. *Log on to the Log Service console*.
2. Click the target project name to go to the corresponding Logstores page. On the Logstores page that appears, click Create.
3. Configure the parameters for the Logstore, and then click OK.

Parameter	Description
Logstore Name	The Logstore name. It must be 3 to 63 characters in length and can contain lowercase letters, digits, hyphens (-), and underscores (_). It must start and end with a lowercase letter or digit. **Note:** - The Logstore name must be unique within the project to which it belongs. - The Logstore name cannot be modified after the Logstore is created.

Parameter	Description
Data Retention Time	The time the collected logs are kept in the Logstore. Unit: days. Valid values: 1 to 365. Logs are deleted if the specified time is exceeded. You can modify the data retention time after the Logstore is created. To modify the data retention time, you can go to the Logstores page, find the target Logstore, and then click Change in the Actions column corresponding to the Logstore. Set Data Retention Time, and then click Change.
Shards	The number of shards for the Logstore. You can create 1 to 10 shards for each Logstore, and create up to 100 shards for each project.

18.2.6 Configure an index

Log Service supports real-time query and analysis of collected log data.

Context

You must enable and configure an index before you use the query and analysis feature.

Note:

- The collected log data can be queried and analyzed only after you enable an index.
- After an index is modified, the new settings are only applicable to the logs that are collected after the modification. The new settings are not applicable to the logs that are collected before the modification.

Procedure

1. *Log on to the Log Service console*.
2. Select a project and click the name of the project.
3. Select a Logstore and click Search in the LogSearch column.
4. Click Enable in the upper-right corner.

   If you have enabled the index before, choose Index Attributes > Configure.

5. On the Search & Analysis page, configure the index.

   You can configure Full Text Index or Field Index. If both are configured, the field index prevails.

   After you configure an index, you can click Delete to the right of the field to delete the configuration.

Category	Parameter	Description
Full text index	Case sensitive	Specifies whether the index is case-sensitive.
	Chinese characters included	Specifies whether to support Chinese word segmentation.
	Delimiter	Specifies the delimiter to separate keywords.
Field Search	Field name	Specifies the log field name.
	Type	Specifies the field type. Valid values: · text · long · double · JSON
	Alias	Specifies the alias of a column.
	Delimiter	Specifies the delimiter to separate keywords.
	Case sensitive	Specifies whether the index is case-sensitive.
	Chinese characters included	Specifies whether to support Chinese word segmentation.

Category	Parameter	Description
	Statistics	Specifies whether to enable statistical analysis for the field. After enabling statistics, you can analyze the field by using statistics statements.

Note:

If your log is a standard NGINX log, you can configure an index in the NGINX Template.

- **If parsed fields are the same as Default Key Name, the same Actual Key Name is set automatically.**
- **If parsed fields are the same as Default Key Name, you can select the corresponding Actual Key Name.**

If you use the NGINX template, the dashboard cyx_logstore-nginx-dashboard is created by default.

6. **Click OK.**

   The index configurations take effect within one minute.

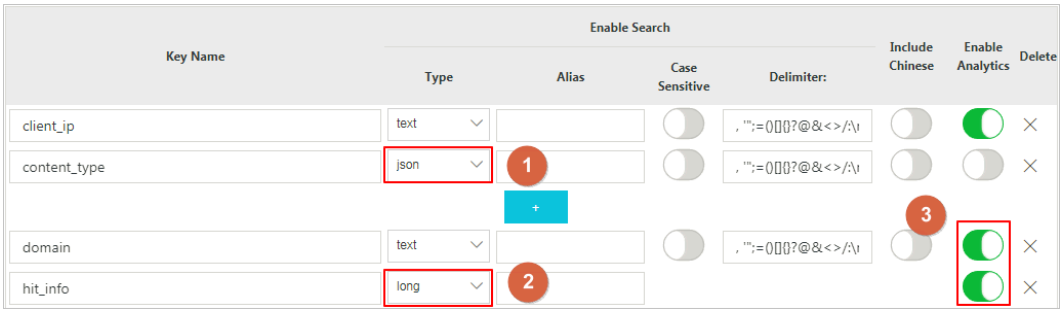
   The following log includes four key values in addition to time.

No.	Key	Type
0	time	-
1	class	text
2	status	long
3	latency	double
4	message	JSON

```
0. time:2018-01-01 12:00:00
  1. class:central-log
  2. status:200
  3. latency:68.75
  4. message:
  {
      "methodName": "getProjectInfo",
      "success": true,
      "remoteAddress": "1.1.1.1:11111",
```

```
        "usedTime": 48,
        "param": {
                "projectName": "ali-log-test-project",
                "requestId": "d3f0c96a-51b0-4166-a850-
f4175dde7323"
        },
        "result": {
            "message": "successful",
            "code": "200",
            "data": {
                "clusterRegion": "ap-southeast-1",
                "ProjectName": "ali-log-test-project",
                "CreateTime": "2017-06-08 20:22:41"
            },
            "success": true
        }
    }
```


The settings are as follows:



Where:

- **① indicates querying all the data of the string and bool types in JSON fields.**
- **② indicates querying data of the long type.**
- **③ indicates SQL analysis of configured fields.**

What's next

After you configure an index, you can retrieve and analyze the collected log data in real time by entering statements in the search box on the query page and clicking Search.

18.2.7 Configure an alert

You can save a saved search as an alert on the query page. Log Service then checks log data at a specified time interval and sends alerts to you when alert conditions are met.

Prerequisites

- **Log data has been collected.**

- Indexes have been enabled and configured. For more information, see *Configure an index*.

Procedure

1. *Log on to the Log Service console*.
2. On the Logstores page, click Search in the LogSearch column.
3. Enter a query statement in the search box.
4. Set a time range for query and click Search.
5. Click Saved as Alarm in the upper-right corner of the page.
6. Configure alert rules and click OK.

Parameter	Description
Alarm Name	The name of the alert rule. · The name can contain lowercase letters, digits, hyphens (-), and underscores (_). · The name must start and end with a lowercase letter or digit. · The name must be 3 to 63 characters in length.
Attribute	
Saved Search Name	The name of the saved search used to configure the alert.
Time Range (minute)	The time range of the data read by the server for each alert check. For example, if the value is set to 1, the server will query the data that is generated within 1 minute before an alert check. You can specify an integer from 1 to 60. Note: The server only samples and processes the first 10 data records generated during the time range when performing an alert check.
Check Interval (min)	The interval at which the server performs alert checks. You can specify an integer from 5 to 1440.

Parameter	Description
Number of Triggers	The number of consecutive alert check triggers. An alert notification is sent when the specified trigger count is reached. You can specify an integer from 1 to 10000. For example, if Check Interval (min) is set to 5 and Number of Triggers is set to 2, the server performs a check every 5 minutes and sends an alert if the results of two consecutive checks meet the alert conditions. The minimum interval between alert notifications is 10 minutes.
Check Condition	
Key	The name of the key used for alerting in the log.
Operator	The comparison operator in the check condition. It can be of the numeric or character type. For more information about comparison operators, see *Table 18-1: Comparison operator*.
Threshold	The comparison value in the check condition. This value is combined with Operator to determine whether saved search results meet the alert conditions.
Action	
ActionType	The method for sending alert notifications. When the configured alert rule is triggered, Log Service sends an alert based on the preset notification method. Alerts can only be sent through the WebHook-Custom method. Notifications are sent to the custom webhook URL through the Post method.
WebHook URL	The URL of webhook. The webhook URL can contain a maximum of 256 characters. It must start with http:// or https://.

Parameter	Description
Content	The content of an alert notification. The content can contain a maximum of 500 characters.

Table 18-1: Comparison operator

Operator	Description	Example
Greater Than	Checks whether a column value is greater than the specified value.	$count > 0
Less Than	Checks whether a column value is smaller than the specified value.	$count<200
Greater Than or Equal to	Checks whether the column value is greater than or equal to the specified value.	$count>=0
Less Than or Equal to	Checks whether the column value is smaller than or equal to the specified value.	$count<=0
Include	Checks whether the specified characters are included.	$project like “admin”
RegEx	Checks whether a string matches the given regular expression.	$project regex match “^/S+$”

Result

You can view the detailed alert records after creating an alert rule.

1. **On the Logstores page, choose LogSearch/Analytics > Alarm from the left-side navigation pane.**
2. **Select an alert rule and click View to view the detailed alert records or the alert status.**

Status:

- **Success: indicates that the rule is executed. You can click Triggering Details to view the trigger conditions.**

- **Failure: indicates that the rule fails to be executed during the query, alert rule matching, or notification phase. You can click Triggering Details to obtain more information about the execution failure.**
  - **Query failed: The query syntax is incorrect.**
  - **Query call failed: Check your network connectivity.**
  - **Failed to call the rule: Check whether rule parameters and response data are in the same format.**

18.2.8 Log consumption

Log Service supports log consumption in a variety of ways.

You can use the following three methods to consume logs that are collected to LogHub of Log Service.

Method	Scenario	Timeliness	Storage duration
Real-time consumption (LogHub)	Stream computing and real-time computing	Real-time (< 10 ms)	365 days
Index query (LogSearch)	Applicable to online query of recent hot data	Real-time	365 days

LogHub

Consumption process

Logs are consumed after being written. Both log consumption and query require the capability of reading logs. The following procedure describes the consumption process of logs in a shard.

1. **Obtain a cursor based on conditions such as time, Begin, and End.**
2. **Read logs based on the cursor and step parameters and return the next cursor.**
3. **Keep moving the cursor to consume logs.**

Consumption methods

Besides the basic API operations, Log Service provides many other methods to consume logs, such as the SDK, Storm spout, Spark Streaming client, Flink connector, consumer library, and Web console.

- **Use the Spark Streaming client to consume logs.**

- **Use the Storm spout to consume logs.**
- **Use the Flink connector to consume logs. The Flink connector consists of Flink consumer and producer.**
- **Use the LogHub consumer library to consume logs. The LogHub consumer library is an advanced consumption mode for LogHub consumers. It provides a lightweight computing framework to implement automatic shard allocation and sequential consumption of logs when multiple LogHub consumers consume Logstores simultaneously.**
- **Use SDKs to consume logs. Log Service provides SDKs in multiple programming languages such as Java and Python that support log consumption API operations. For more information about the SDKs, see** *Log Service SDK reference* **.**

LogSearch

- **Query logs in the Log Service console.**
- **Query logs by using the SDKs or API operations of Log Service. Log Service provides HTTP-based RESTful API operations. The API supports comprehensive log queries. For more information, see** *Log Service API reference* **.**

18.3 Project management

18.3.1 Project

A project is the resource management unit of Log Service and is used to isolate and control resources. You can use a project to manage all the logs and related log sources of an application. A project is used to manage Logstores of a user and machine configurations for log collection. A project also serves as the portal for a user to access the resources of Log Service.

Projects provide the following features:

- **Help you to organize and manage different Logstores. You can use Log Service to collect and store the logs of different projects, services, or environments. You can classify different logs for management in different projects to facilitate subsequent log consumption, exporting, or indexing. In addition, projects are the carriers for log access control.**
- **Provide you with a portal to access Log Service resources. Log Service allocates an exclusive access portal to each project. The access portal allows you to write, read, and manage logs through the network.**

You can use the Log Service console to perform the following project-related operations:

- *Create a project*
- *Manage a project*
- *Delete a project*

18.3.2 Import a project

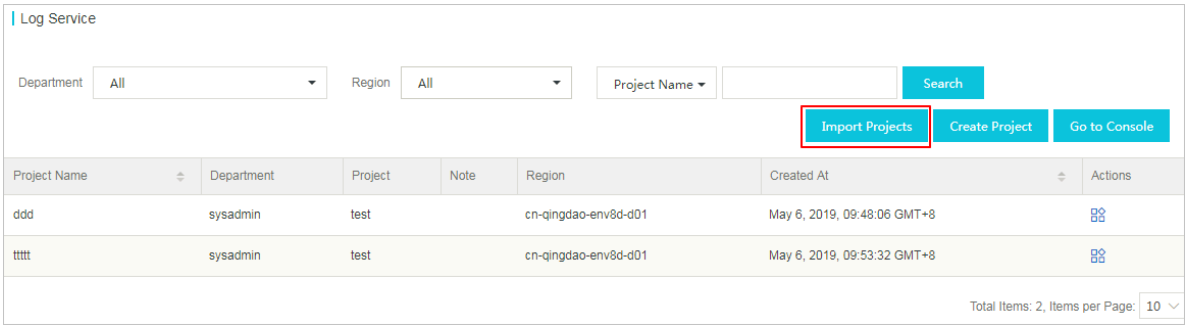
You can update the project list on the Log Service homepage of the Apsara Stack console by using the project import feature. After the update, the projects that you have created in the Log Service console are displayed in the project list on the Log Service homepage of the Apsara Stack console.

Context

By default, the projects that you have created in the Log Service console are only displayed in the project list of the console. You can use the project import feature to import the created projects to the project list on the Log Service homepage of the Apsara Stack console.

Procedure

1. *Log on to the Apsara Stack console and go to the Log Service homepage*.
2. Click Import Projects in the upper-right corner.



3. Select a department and click OK.

Result

After projects are imported, the project list on the Log Service homepage of the Apsara Stack console is updated.

What's next

Change project ownership

18.3.3 Change project ownership

After creating a project, you can change the department and project to which the project belongs.

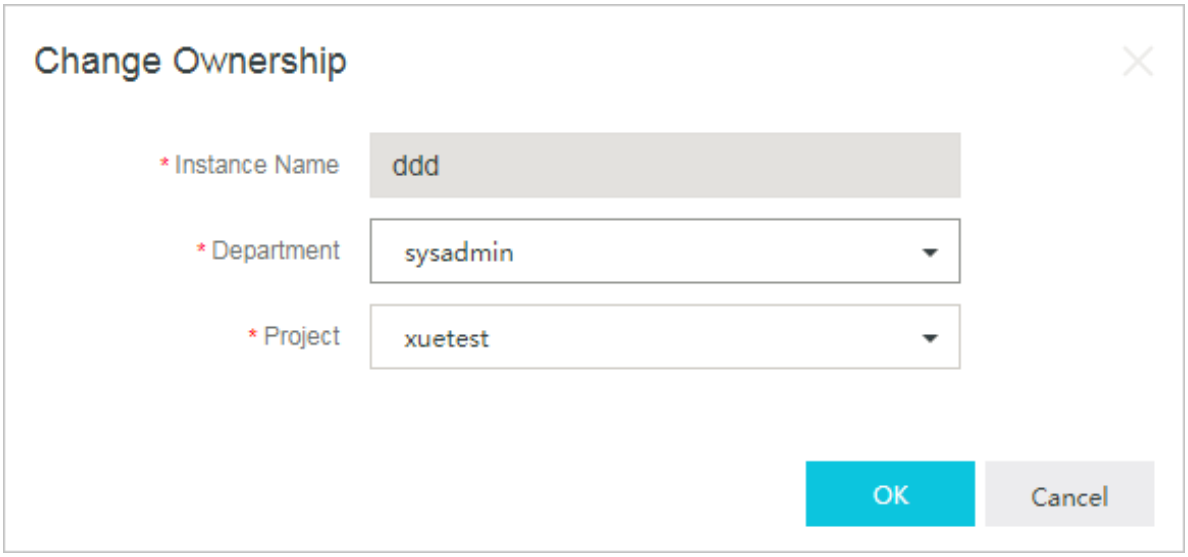
Context

You can migrate the project to another department and project by changing the ownership of the project.

> **Note:**
> **If your project is created in the Log Service console, it is automatically created in a level-1 department and does not belong to any project. In this case, you must *Import a project* to the Apsara Stack console and migrate your project to another project by changing its ownership.**

Procedure

1. *Log on to the Apsara Stack console and go to the Log Service homepage*.
2. Click the icon in the Actions column corresponding to the specified project name.
3. Choose Change Ownership from the shortcut menu.
4. Select a different Department and Project for the project.



5. Click OK.

18.3.4 Modify a project comment

You can add project comments to facilitate project management.

Context

A project is the resource management unit of Log Service and is used to manage Logstores and machines for log collection. A Logstore is the log storage unit of Log Service and is used to store a specific type of logs. A project can collect multiple types of logs, such as access logs of front-end Web servers and application logs generated by back-end applications. You can create separate Logstores for a project and write different types of logs to different Logstores.

Procedure

1. *Log on to the Log Service console*.
2. On the Projects page, click Change Comment in the Actions column corresponding to the specified project.
3. In the dialog box that appears, modify the project comment and click OK.

18.3.5 Delete a project

In some cases such as disabling Log Service or destroying all logs in a project, you may need to delete the project. You can delete the entire project in the Log Service console.

Prerequisites

If the project is created in the Log Service console, the project is not displayed on the Log Service homepage of the Apsara Stack console by default. You must *Import a project* **before you can delete a project.**

Context

After you delete a project, all its logs and configurations are permanently deleted and cannot be recovered. Therefore, proceed with caution when deleting a project to avoid data loss.

Procedure

1. *Log on to the Apsara Stack console and go to the Log Service homepage*.
2. Click the  icon in the Actions column corresponding to the specified project.
3. In the dialog box that appears, click Delete Project.

4. A dialog box is displayed, prompting you to confirm whether to delete the project. If yes, click OK.

18.4 Logstore management

18.4.1 Logstores

A Logstore is the unit used in Log Service for log data collection, storage, and query. Each Logstore can belong to only one project, and multiple Logstores can be created for a single project.

You can create multiple Logstores for a project as needed. An independent Logstore is created for each type of log in an application. For example, you have a game called `big-game`, and it stores three types of logs on the server: operation_log, application_log, and access_log. You can create a project named `big-game`, and create three Logstores under this project to collect, store, and query the three types of logs. Whether writing or querying logs, you must specify a Logstore for the actual operation.

Logstores provide the following features:

- Log collection: Logstores support real-time logging.
- Log storage: Logstores support real-time consumption.
- Index creation: Logstores support real-time log query.

You can perform the following Logstore operations in the Log Service console:

- *Create a Logstore*
- *Modify Logstore configurations*
- *Delete a Logstore*

18.4.2 Modify Logstore configurations

After a Logstore is created, you can modify the Logstore configurations as needed.

Procedure

1. *Log on to the Log Service console*.
2. Find the target project and click the project name.

   You can view Logstores that have been created for the project.
3. On the Logstores page, find the target Logstore and click Modify in the Actions column corresponding to the Logstore.

4. In the dialog box that appears, modify the Logstore configurations and click OK.

   For more information about how to adjust shards, see *Split shards*.

Parameter	Description
Logstore Name	The Logstore name. It must be 3 to 63 characters in length and can contain lowercase letters, digits, hyphens (-), and underscores (_). It must start and end with a lowercase letter or digit. **Note:** · The Logstore name must be unique within the project to which it belongs. · The Logstore name cannot be modified after the Logstore is created.
Data Retention Time	The time the collected logs are kept in the Logstore. Unit: days. Valid values: 1 to 365. Logs are deleted if the specified time is exceeded. You can modify the data retention time after the Logstore is created. To modify the data retention time, you can go to the Logstores page, find the target Logstore, and then click Change in the Actions column corresponding to the Logstore. Set Data Retention Time, and then click Change.
Shards	The number of shards for the Logstore. You can create 1 to 10 shards for each Logstore, and create up to 100 shards for each project.

18.4.3 Delete a Logstore

You may need to delete a Logstore that you no longer use. You can delete a Logstore in the Log Service console.

Precautions

- After a Logstore is deleted, its log data will be lost permanently and cannot be recovered, and a Logstore with the same name cannot be created. Proceed with caution.
- Before deleting a Logstore, you must delete all its Logtail configurations.

Procedure

1. *Log on to the Log Service console*.
2. Select a project and click the name of the project to go to the Logstores page.
3. On the Logstores page, select the Logstore to be deleted and click Delete.
4. In the dialog box that appear, click OK.

18.5 Shard management

18.5.1 Manage shards

Logstore read/write logs must be saved in a shard. Each Logstore has several shards. Each shard is represented by a non-overlapping, left-closed, and right-open interval of MD5 values. The range of a Logstore is represented by the entire range of MD5 values of the shards in the Logstore.

Range

You must specify the number of shards when creating a Logstore. The entire MD5 value range is automatically and evenly divided based on the specified number of shards. Each shard has a range, which can be expressed as MD5 values and must be within the following value range: [00000000000000000000000000000000,ffffffffffffffffffffffffffffffff).

All of the shard ranges are left-closed and right-open intervals, which involve the following keys:

- BeginKey: specifies the start of a shard. The value of this key is included in the shard range.
- EndKey: specifies the end of a shard. The value of this key is excluded from the shard range.

The shard range allows you to use hash keys to write logs to specific shards, and to identify shards to split or merge. When reading data from a shard, you must specify the corresponding shard. When writing data to a shard, you can use the load balancing or hash key mode. In load balancing mode, each data packet is randomly written to any available shard. In hash key mode, data is written to the shard whose range includes the specified key value.

For example, a Logstore has four shards and the MD5 value range of this Logstore is [00,FF). *Table 18-2: Shard example* describes the range of each shard.

Table 18-2: Shard example

Shard	Range
Shard0	[00,40)
Shard1	[40,80)
Shard2	[80,C0)
Shard3	[C0,FF)

If you specify the MD5 key value as 5F when writing logs in hash key mode, the log data is written to Shard1 because Shard1 contains the MD5 key value 5F. If you specify the MD5 key value as 8C, the log data is written to Shard2 because Shard2 contains the MD5 key value 8C.

Read/write capacities

Each shard provides certain read/write capacities. We recommend that you plan the number of shards based on the actual data traffic. If the data traffic exceeds the read/write capacities, increase the number of shards by splitting shards to achieve greater read/write capacities. If the data traffic is far less than the maximum read /write capacities of shards, reduce the number of shards by merging the shards to save costs.

For example, you have two shards in the readwrite state and can write at 10 MB/s at maximum. If you write data at 14 MB/s in real time, we recommend that you split a shard in the readwrite state to reach three. If you write data at only 3 MB/s, we recommend that you merge these two shards because one shard is sufficient.

Note:

- **If the API constantly reports 403 and 500 errors during writing, check Log Service monitoring metrics and determine whether to increase the number of shards.**
- **For read/write operations that exceed the service capacities of shards, the system attempts to provide the needed services, but the service quality cannot be guaranteed.**

Shard status

Shard status includes:

- **readwrite: Supports reading and writing data.**
- **readonly: Only supports reading data.**

When a shard is created, it is in the readwrite state. Split or merge operations change the state of original shards to readonly and generate new shards in the readwrite state. The state of shards does not affect the performance of reading data. Data can be written to shards in the readwrite state, but not to shards in the readonly state.

When splitting a shard, you must specify the ID of a shard in the readwrite state and an MD5 value. The MD5 value must be greater than the BeginKey value of the shard and less than the EndKey value of the shard. Split operations can split two other shards from one. The number of shards is increased by two after each split. After a shard is split, the state of the original shard is changed from readwrite to readonly. Data can still be consumed, but new data cannot be written to the original shard. The two new shards are in the readwrite state and arranged behind the original shard. The MD5 range of these two shards covers the range of the original shard.

When merging shards, you must specify a shard in the readwrite state. Make sure the specified shard is not the last shard in the readwrite state. Log Service automatically finds the adjacent shard at the right of the specified shard and merges these two shards. After the merge, the specified shard and the adjacent shard are in the readonly state. Data can still be consumed, but new data cannot be written to the merged shards. A new shard in the readwrite state is generated, and its MD5 value range covers the total range of the original two shards.

In the Log Service console, you can perform the following shard-related operations:

- **Scale out shards**
- **Scale in shards**

Note:

The lifecycle (or data retention period) of a Logstore ranges from 1 to 365 days. Shards and their logs will be automatically deleted when this period expires.

18.5.2 Split shards

When creating a Logstore, you can select the number of shards based on the volume and generation speed of your logs. You can also change the number of shards by splitting or merging shards when modifying a Logstore.

Context

Read/write logs in a Logstore must be saved in a shard. Each Logstore has several shards. When creating a Logstore, you must specify the number of shards. After the Logstore is created, you can split or merge shards to increase or reduce shards.

Each shard can write data at 5 MB/s and read data at 10 MB/s. When the data traffic exceeds the service capability of the shards, we recommend that you add shards immediately. A shard can be split to scale out its service capacity.

When splitting a shard, you must specify the ID of a shard in the read/write state and an MD5. The MD5 must be greater than the shard BeginKey and smaller than the shard EndKey.

A split operation can split one shard to generate two new shards. After a split operation, two more shards are added. After a shard is split, the status of the original shard is changed from read/write to read-only. Data can still be consumed , but new data cannot be written. The two new shards are in the read/write state and located behind the original shard. The MD5 ranges of these new shards are the range of the original shard.

Procedure

1. *Log on to the Log Service console*.
2. **Select a project and click the name of the project.**
3. **On the Logstores page, select a Logstore and click Change in the Actions column.**
4. **Select the shard to be split, and click Split.**
5. **Click OK and close the dialog box.**

   After the split, the status of the original shard is changed to read-only. The MD5 ranges of the two new shards are the range of the original shard.

18.5.3 Merge shards

Shards can be merged to scale in their service capacity. The merge operation merges the ranges of a specified shard and its adjacent shard on the right, and

assigns the combined range to a new shard in the read/write state. The original shards then enter the read-only state.

When merging shards, you must specify a shard in the read/write state. Make sure that the specified shard is not the last shard in the read/write state. The server automatically finds the adjacent shard at the right of the specified shard and merges these two shards. After the merge, the merged shards enter the read-only state. Data can still be consumed, but new data cannot be written. A shard in the read/write state is generated, and its MD5 range is the MD5 range of the original shards.

Procedure

1. *Log on to the Log Service console*.
2. Select a project and click the name of the project.
3. On the Logstores page, select a Logstore and click Change in the Actions column.
4. Select the shard to be merged and click Merge on the right. Then, close the dialog box.

   After the merge, the specified shard and its adjacent shard on the right enter the read-only state. The data in these two shards can still be consumed. A shard in the read/write state is generated, and its MD5 range is the total range of the original shards.

18.6 Data collection

18.6.1 Data collection overview

LogHub supports a variety of methods for lossless log collection such as clients, Web pages, protocols, SDKs, and APIs. You can select a collection method for your data sources as needed.

18.6.2 Collect NGINX access logs

Log Service allows you to query and analyze real-time logs, and saves the analysis results to the dashboard. These features greatly decrease the complexity in analyzing NGINX access logs and facilitate the collection of website access statistics.

Context

Many webmasters use NGINX as the server to build websites. When analyzing the website access data, they must perform statistical analysis on NGINX access logs to obtain data such as the page views and access time periods of the website. In traditional methods such as CNZZ, a JavaScript script is inserted in the front-end page and is triggered when a user accesses the website. However, this method can only record access requests. Stream computing and offline statistics and analysis can also be used to analyze NGINX access logs. However, this method requires that an environment be established, and is subject to imbalance between timeliness and analytical flexibility.

Log Service allows you to query and analyze real-time logs, and saves the analysis results to the dashboard. These features greatly decrease the complexity in analyzing NGINX access logs and facilitate the collection of website access statistics. This topic describes how to implement log analysis in the NGINX access log scenario.

Log formats

We recommend that you use the following log_format configuration to better match the analysis scenario:

```
    log_format  main  '$remote_addr - $remote_user [$time_local] "$
request" $http_host '
                '$status $request_length $body_bytes_sent "$http_refer
er" '
                '"$http_user_agent"  $request_time $upstream_r
esponse_time';
```

The following table describes the meaning of each field.

Field	Description
remote_addr	The IP address of the client.
remote_user	The username of the client.
time_local	The server time.
request	The request content, including the method name, address, and HTTP.
http_host	The HTTP address used by the user request.
status	The returned HTTP status code.
request_length	The size of the request packet.

Field	Description
body_bytes_sent	The size of the response packet.
http_referer	The referer.
http_user_agent	The name of the client.
request_time	The overall request latency.
upstream_response_time	The processing latency of upstream services.

Procedure

1. *Log on to the Log Service console*.
2. Click a project name.
3. On the Logstores page, click the icon in the Data Import Wizard column corresponding to a Logstore.

4. **Configure the data source.**

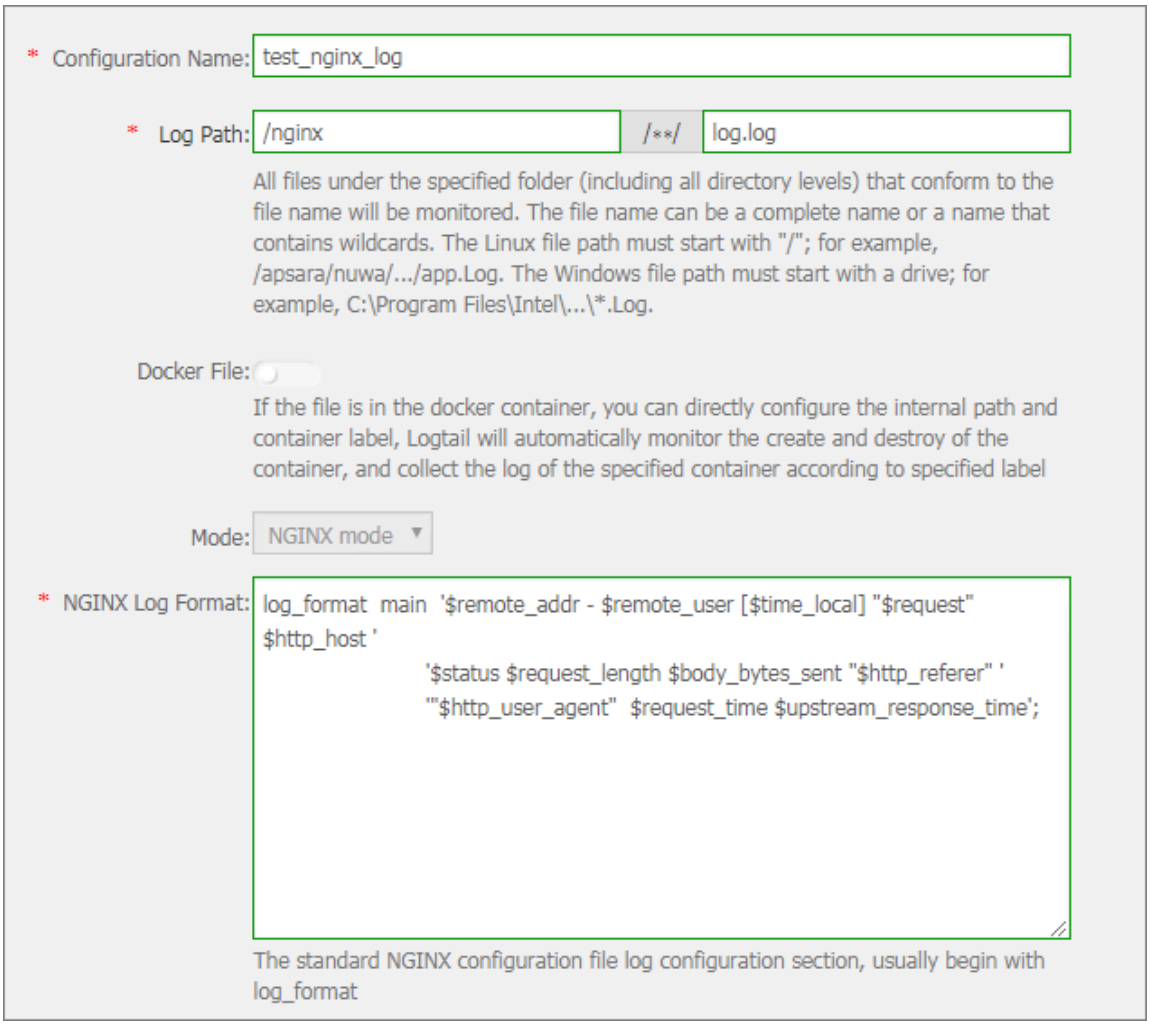
   a) **Select NGINX Access Log.**

   b) **Set Configuration Name.**

   c) **Set Log Path.**

   d) **Set NGINX Log Format.**

   Enter your `log_format` information in the NGINX Log Format field.



   e) **Confirm NGINX Key.**

   Log Service automatically extracts the corresponding keys.

Note:

`$request` **is extracted as two keys:** `request_method` **and** `request_uri`**.**

f) Optional: Set Advanced Options.

Parameter	Description
Local Cache	This parameter specifies whether to turn on Local Cache. When Log Service is unavailable, logs can be cached to a local directory and then uploaded after Log Service recovers. By default, up to 1 GB logs can be cached.
Upload Raw Log	This parameter specifies whether to upload the raw log. If you enable this function, the raw log content is uploaded as the __raw__ field with the parsed log content.

Parameter	Description
Topic Generation Mode	· Null - Do not generate topic: By default, this mode is selected. In this mode, the topic is set to a null string and you can query logs without entering the topic. · Machine Group Topic Attributes: This mode differentiates log data generated in different front-end servers. · File Path RegEx: When this mode is selected, you must configure Custom RegEx below to extract a part of the path as the topic. This mode differentiates log data generated by different users or instances.
Custom RegEx	If you select File Path RegEx, you must enter a custom regular expression.
Log File Encoding	· utf8: specifies UTF-8 encoding. · gbk: specifies GBK encoding.
Maximum Directory Monitoring Depth	This parameter specifies the maximum level of the directories that are monitored when logs are collected from the log source. The value ranges from 0 to 1000 . The value 0 indicates that only the directory at the current level is monitored.
Timeout	If a log file is not updated within the specified period of time, the system considers that the file has timed out. The following Timeout options are available: · Never: All log files are continuously monitored and never time out. · 30 Minute Timeout: If a log file is not updated in 30 minutes, the system considers that the log file has timed out and no longer monitors the file.

Parameter	Description
Filter Configuration	Only logs that completely meet the filtering conditions are collected. For example, `Key:level Regex:WARNING|ERROR` indicates that only WARNING and ERROR logs are collected. You can also filter logs that do not conform to a condition. For example, `Key:level Regex:^(?!.*(INFO|DEBUG))` indicates that INFO and DEBUG logs are not collected. For more information about similar examples, see *regex-exclude-word* and *regex-exclude-pattern*.

g) After completing the settings, click Next.

5. Select the target machine group and click Apply to Machine Group.

   You must create a machine group before performing this operation.

6. **Optional: Configure search, analysis, and visualization.**

   When the heartbeat status of the machine group is normal, you can click Preview to view the collected log data.

   Log Service provides predefined keys for analysis. You can select the actual keys that are generated based on the preview data and map them with the default keys .

Click Next. Log Service configures the index attributes for you and creates the `nginx-dashboard` dashboard for analysis.

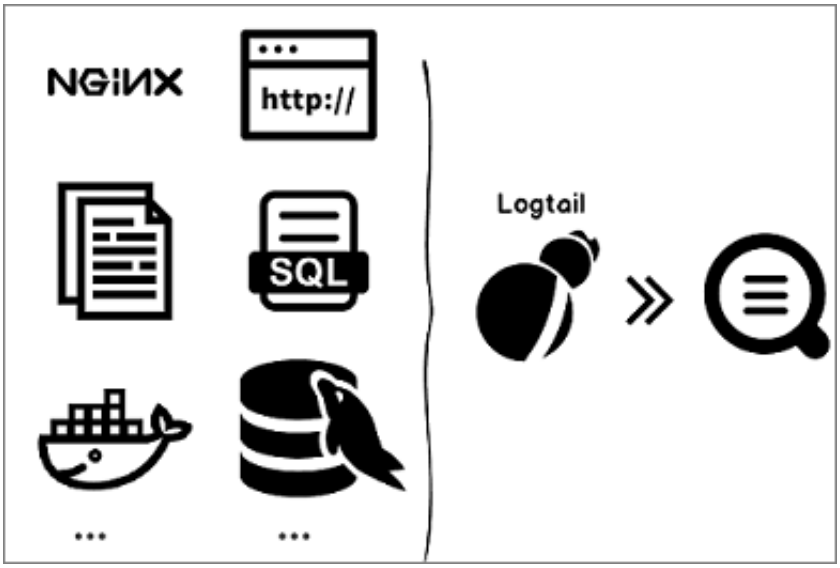
18.7 Collection by Logtail

18.7.1 Overview

18.7.1.1 Logtail overview

Log Service provides a log collection agent: Logtail. Logtail allows you to collect logs from servers such as ECS instances in the console in real time. Currently,

Apsara Stack Log Service only supports Logtail on a Linux server. To collect logs from a Windows server, use Logstash.



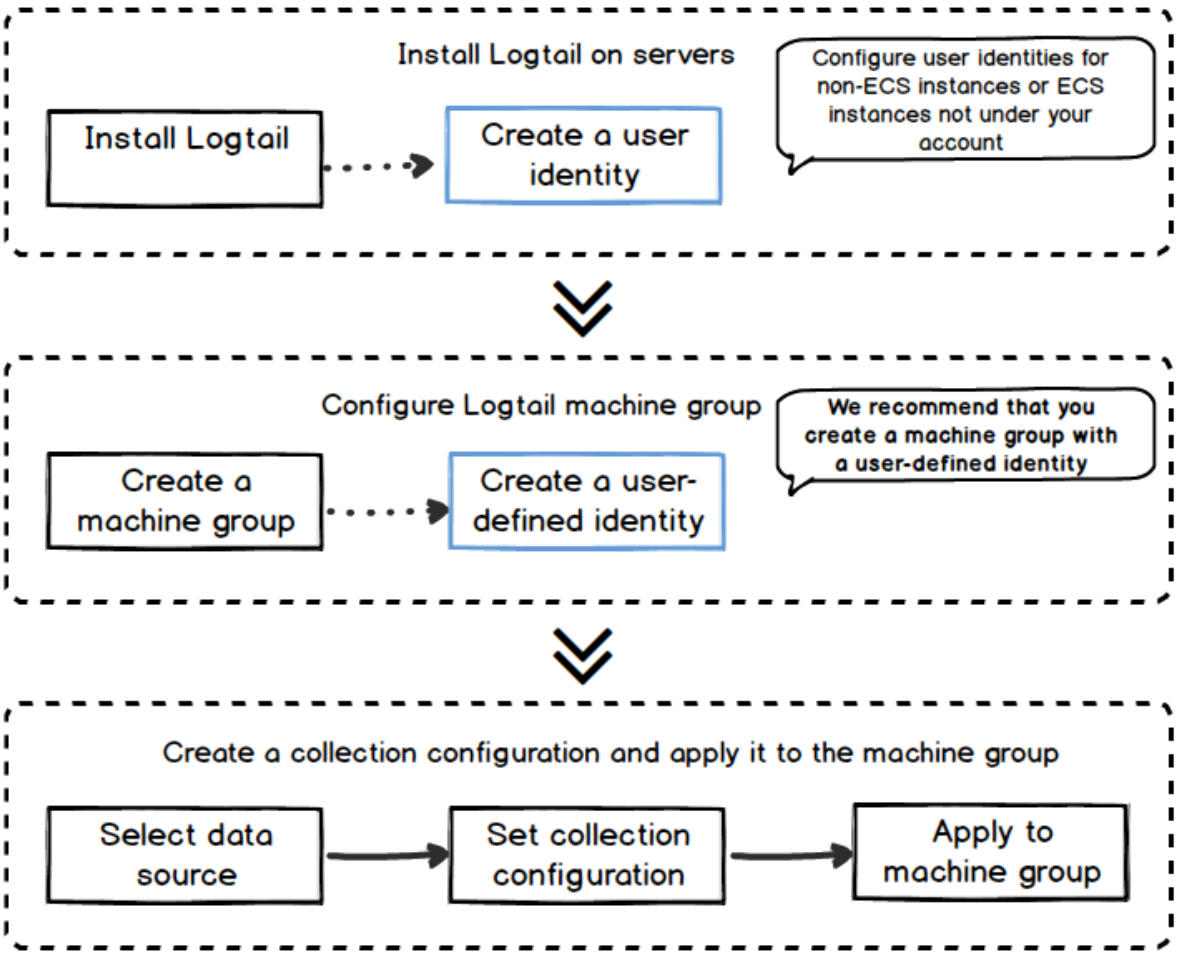
Benefits

- **Provides non-intrusive log collection based on log files. You do not need to modify any application code, and log collection does not affect the operating logic of your applications.**
- **Supports exception handling in a stable manner during the log collection process. If network and Log Service exceptions occur or the data write bandwidth temporarily exceeds the reserved write bandwidth, Logtail actively retries and caches data locally to guarantee data security.**
- **Supports central management based on Log Service. After installing Logtail, you only need to configure the servers from which you want to collect logs and the collection method in Log Service, eliminating the need to log on to servers one by one. For more information about how to install Logtail, see *Install Logtail in Linux*.**
- **Provides a comprehensive self-protection mechanism. To ensure that the collection agent running on servers does not have obvious impacts on the performance of services, Logtail provides a protective mechanism and strictly limits its use of CPU, memory, and network resources.**

Processing capabilities and limits

See *Limits*.

Configuration process



Perform the following steps to use Logtail to collect logs from servers:

1. **Install Logtail. Install Logtail on servers from which you want to collect logs. For more information, see** *Install Logtail in Linux*.
2. *Create a machine group*. **Log Service uses machine groups to manage all servers from which you want to collect logs with Logtail. Log Service allows you to define machine groups by using IP addresses or user-defined identifiers. You can create a machine group as instructed when applying Logtail configurations to machine groups.**
3. **Create Logtail configurations and apply them to a machine group. You can use Data Import Wizard to create Logtail configurations for collecting** *text files* **and** *syslog data*, **and to apply the Logtail configurations to a machine group.**

After the preceding process is complete, logs of a specific type on the ECS instances are collected and sent to the selected Logstore. Historical logs are not collected. You can use the Log Service console, SDKs, or APIs to query these logs. Log Service

allows you to view the Logtail collection status on each ECS instance, for example, whether the collection is normal and whether any error occurs.

For more information about the complete Logtail configuration operations on the Log Service console, see *Collect logs by using Logtail*.

Docker

- **Container Service: For more information, see Integrated Log Service in** *Container Service User Guide* .
- **User-created Docker for ECS instances and IDCs: Mount the log directories from containers to the host server.**
  - *Install Logtail in Linux*.
  - **Mount the log directories from containers to the host server.**
    - **Method 1: Use commands. For example, if the directory on the host server is** `/log/webapp` **and the directory in a container is** `/opt/webapp/log`**, run the following command:**

      ```
      docker run -d -P --name web -v /src/webapp:/opt/webapp training
      /webapp python app.py
      ```

    - **Method 2: Use the orchestration template.**

Note:

We recommend that you modify the Logtail startup parameters, change the checkpoint storage path of Logtail, and mount the log directories to the host server. This method prevents repeated collection due to loss of checkpoint information when containers are released.

Terms

- **Machine group: A machine group contains one or more servers from which logs of a specific type are collected. You can apply Logtail configurations to a machine group to enable Log Service to collect logs from all servers in the machine group by using the same Logtail configurations. The Log Service console provides a convenient way to manage machine groups, including operations to create, delete, add, and remove servers.**

- **Logtail client: Logtail is the agent that collects logs from the servers running Logtail. For more information, see *Install Logtail in Linux*. After installing Logtail on a server, create a Logtail configuration and apply it to a machine group.**
  - **In Linux, Logtail is installed in the `/usr/local/ilogtail` directory and starts two separate processes whose names begin with ilogtail: collection process and daemon. The program running log is `/usr/local/ilogtail/ilogtail.LOG` .**
- **Logtail configuration: Logtail configuration is a collection of policies to collect logs by using Logtail. By configuring Logtail parameters such as data source and collection mode, you can customize the log collection policy for all the servers in a machine group. Logtail configurations describe how to collect specific types of logs from servers, parse the collected logs, and send the logs to the specified Logstore of Log Service. You can add a Logtail configuration for a Logstore in the console to enable the Logstore to receive logs collected by using this Logtail configuration.**

Features

Logtail provides the following features:

- **Real-time log collection: Logtail dynamically monitors log files, reads them in real time, and parses incremental logs. Generally, logs are sent to Log Service within 3 seconds after they are generated.**

  Note:
  Logtail does not support the collection of historical data. If logs are read 5 minutes or later after they are generated, the logs will be discarded.
- **Automatic log rotation processing: Many applications rotate log files based on the file size or generation date. During the rotation process, original log files are renamed and new empty log files are created. For example, the monitored `app.LOG` is rotated, generating files such as `app.LOG.1` and `app.LOG.2`. You can specify the file to which collected logs are written, such as `app.LOG`. Logtail automatically detects the log rotation process and ensures that no logs are lost during this process.**

  Note:

However, if log files are rotated multiple times within a few seconds, data loss may occur.

- **Automatic handling of collection exceptions: When data transmission fails because of exceptions such as Log Service errors, network errors, and exceeded quotas, Logtail actively retries based on the specific scenario. If the retry fails, Logtail writes the data to the local cache and then automatically resends the data later.**

   **Note:**
  The local cache is located in the disk of your server. If the data cached locally is not received by Log Service within 24 hours, the data is discarded and deleted from the local cache.

- **Flexible configuration of collection policies: You can use Logtail configurations to specify how logs are collected from an ECS instance. Specifically, you can select log directories and files, which support exact match or fuzzy match with wildcards, based on actual scenarios. You can customize the extraction method of collected logs and the names of extracted fields. Log Service allows you to extract logs by using regular expressions. The log data models of Log Service require that each log must have a precise timestamp. Logtail provides custom log time formats, allowing you to extract the required timestamp information from log data of different formats.**
- **Automatic synchronization of collection configurations: After you create or update a configuration in the Log Service console, Logtail automatically receives the configuration and brings the configuration into effect within 3 minutes. No collected data is lost when a configuration is being updated.**
- **Automatic client upgrade: After you manually install Logtail on a server, Log Service manages the automatic upgrades of Logtail without manual intervention. No log data is lost during the Logtail upgrade process.**
- **Status monitoring: To avoid consuming too many resources and thus affecting your services, the Logtail client monitors its consumption of CPU and memory in real time. The Logtail client automatically restarts when its resource usage exceeds the limit to avoid affecting other services on the server. The**

Logtail client proactively limits network traffic to avoid excessive bandwidth consumption.

> **Note:**
>
> - Log data may be lost when the Logtail client restarts.
> - If the Logtail client exits because of an exception in its processing logic, the corresponding protection mechanism is triggered and the Logtail client is restarted and continues to collect logs. However, log data generated before the restart may be lost.

- Transferred data signature: To prevent data tampering during data transfer, the Logtail client proactively obtains your AccessKey pair to sign all log data packets before sending them.

> **Note:**
> The Logtail client uses an HTTPS channel to obtain the AccessKey pair. This ensures the security of the AccessKey pair.

18.7.1.2 Logtail collection principles

The Logtail client performs the following steps to collect server logs: listen for files, read files, process logs, filter logs, aggregate logs, and send data.

After you install a Logtail client on a server and add a log collection configuration to the Logtail client, Logtail starts to collect logs to Log Service. Logtail collects logs in the following steps:

1. Listen for files
2. Read files
3. Process logs
4. Filter logs
5. Aggregate logs
6. Send logs

> **Note:**
> After the Logtail collection configuration is applied to a machine group, unmodified logs on the servers in the machine group are considered as historical files. In the normal running mode, Logtail does not collect historical files.

Listen for files

After the Logtail client is installed on the server and the Logtail collection configuration is added based on the data source, the Logtail collection configurat ion is delivered from the server to Logtail in real time. Logtail starts to listen for files based on the collection configuration.

1. **Logtail scans log directories and files that comply with the specified file name rules based on the configured log path and maximum monitoring directory depth.**

   To ensure the timeliness and stability of log collection, Logtail listens for the registration events of collected directories and performs periodic polling.
2. **If log files in the specified directory that match the rules are not modified after the configuration is applied, Logtail does not collect these log files. If some log files are modified, Logtail triggers the collection process and reads these files.**

Read files

After determining that a log file has been updated, Logtail reads the log file.

1. **Logtail checks the size of a log file that is read for the first time.**
   - **If the file size is less than 1 MB, Logtail reads from the beginning of the file.**
   - **If the file size is larger than 1 MB, Logtail reads the last 1 MB of data in the file.**
2. **If Logtail has read the file, Logtail continues to read it from the last checkpoint.**
3. **Logtail can read up to 512 KB of data at a time. Therefore, you need to limit the log size to 512 KB.**

Process logs

After reading a log, Logtail divides the log into lines, parses the log, and verifies the time field of the log.

1. **Divide into lines:**

   If regular expression at the beginning of the line is specified in the Logtail configuration, every time Logtail reads log data, it divides the log into lines based on the configuration. If the beginning of the line is not set, each data block is processed as a log.

2. **Parse the log:**

   Logtail parses each log based on the collection configuration, such as regular expressions, delimiters, and JSON.

   Note:
   A complicated regular expression may lead to high CPU utilization. Therefore, we recommend that you use an efficient regular expression.

3. **Handle parsing failures:**

   You can determine how to handle parsing failures based on whether to enable the Discard Logs with Parsing Failures feature in the Logtail collection configurat ion.

   - **If this feature is enabled, Logtail discards logs that fail to be parsed. and reports errors.**
   - **If this feature is disabled, Logtail uploads the original logs that fail to be parsed with the Key set to raw_log and the Value set to the log content.**

4. **Set the time field of a log:**

   - **The log time is the current parsing time if the time field is not specified.**
   - **If the time field is specified,**
     - **The log time is extracted from the parsed log fields when the difference between the log time and the current time is less than 12 hours.**
     - **The log is discarded and an error is reported when the difference between the log time and the current time is greater than 12 hours.**

Filter logs

After processing logs, Logtail filters them based on the filtering settings of the Logtail collection configuration.

- **If the filtering settings are not configured, Logtail skips to the next step without filtering any logs.**

- **If the filtering settings are configured, Logtail traverses and verifies all the fields of each log.**
  - **Logs that comply with the filtering settings: Logtail collects the logs that contain all the fields configured by the filter and comply with the settings.**
  - **Logs that do not comply with the filtering settings: Logtail does not collect the logs that do not comply with the filtering settings.**

Aggregate logs

After logs are filtered based on the log filtering settings, Logtail sends the logs that comply with these settings to Log Service. To reduce the number of requests , Logtail caches the processed and filtered logs for a period of time. Then, Logtail aggregates and packages these logs before sending them to Log Service.

If one of the following conditions is met during caching, logs are immediately packaged and sent to Log Service.

- **Log aggregation lasts more than three seconds.**
- **The number of aggregated logs exceeds 4,096.**
- **The total size of aggregated logs exceeds 512 KB.**

Send logs

Logtail aggregates and sends the collected logs to Log Service. You can set the `max_bytes_per_sec` and `send_request_concurrency` parameters of the startup parameter configuration to adjust the sending rate and maximum concurrency of log data. Logtail can keep the sending rate and concurrency below the configured thresholds.

If data sending fails, Logtail retries or stops sending based on error messages.

Error code	Description	Handling method
Error 401	The Logtail client is not authorized to collect data.	Logtail discards log packets.
Error 404	The specified project or Logstore does not exist in the Logtail collection configuration .	Logtail discards log packets.
Error 403	The shard quota is exhausted.	Logtail waits for three seconds and retries.

Error code	Description	Handling method
Error 500	A server exception has occurred .	Logtail waits for three seconds and retries.
Network timeout	A network connection error has occurred.	Logtail waits for three seconds and retries.

18.7.2 Installation

18.7.2.1 Install Logtail in Linux

You must install Logtail before using it to collect logs. Logtail can only be installed on Linux servers.

Supported systems:

The following versions of Linux x86 (64-bit) servers are supported:

- **Aliyun Linux**
- **Ubuntu**
- **Debian**
- **CentOS**
- **OpenSUSE**

Procedure

1. **Download the Logtail installation script.**

   Run the following command to download Logtail:

   ```
   logtail.your Log Service endpoint/logtail.sh
   ```

2. **Execute the installation script.**

   Start the shell terminal and run the following command as an administrator to install Logtail:

   ```
   sh logtail.sh
   ```

   Note:
   Logtail installation overwrites any existing Logtail installation. If you have installed Logtail, the installer uninstalls and deletes the `/usr/local/ilogtail` **directory before installing Logtail.**

What's next

View the running Logtail version

The following example shows that the version of running Logtail is 0.9.4:

```
$ls /usr/local/ilogtail/ilogtail -lh
lrwxrwxrwx 1 root root 34 Nov  3 12:00 /usr/local/ilogtail/ilogtail -
> /usr/local/ilogtail/ilogtail_0.9.4
```

Uninstall Logtail

See Install Logtail to download *logtail.sh*. Run the following command as an administrator in shell:

```
wget http://{sls data endpoint}/logtail.sh
chmod 755 logtail.sh
sh logtail.sh uninstall
```

18.7.2.2 Configure startup parameters

This topic describes how to configure Logtail startup parameters. You can use this topic as a reference for parameter configuration.

Context

The configuration of Logtail startup parameters is applicable to the following scenarios:

- **Excessive memory space is occupied due to a large number of log files to be collected. The metadata of each file must be maintained in memory, including the file signature, collection location, and file name.**
- **CPU utilization is high due to heavy log data traffic.**
- **A high volume of log data leads to heavy traffic sent to Log Service.**
- **Syslog and TCP data streams need to be collected.**

Startup configuration

- **File path**

  ```
  /usr/local/ilogtail/ilogtail_config.json
  ```

- **File format**

  JSON

- **File sample (only partial configuration items are provided)**

  ```
  {
      ...
      "cpu_usage_limit" : 0.4,
      "mem_usage_limit" : 100,
      "max_bytes_per_sec" : 2097152,
  ```

```
    "process_thread_count" : 1,
    "send_request_concurrency" : 4,
    "streamlog_open" : false,
    "streamlog_pool_size_in_mb" : 50,
    "streamlog_recv_size_each_call" : 1024,
    "streamlog_formats":[],
    "streamlog_tcp_port" : 11111,
    "buffer_file_num" : 25,
    "buffer_file_size" : 20971520,
    "buffer_file_path" : "",
    ...
}
```

Common configuration parameters

Parameter	Value	Description
cpu_usage_limit	The CPU utilization threshold. The value is of the double type and calculated per core.	For example, the value 0.4 indicates that the CPU utilization of Logtail is limited to 40% of a single-core CPU. Logtail restarts automatically when the threshold is exceeded. In most cases, the processing capability of a single-core CPU is about 24 MB/s in simple mode and about 12 MB/s in full regex mode.
mem_usage_limit	The usage threshold of resident memory. The value is of the int type and measured in MB.	For example, the value 100 indicates the memory usage of Logtail is limited to 100 MB. Logtail restarts automatically when the threshold is exceeded. If you need to collect more than 1,000 distinct files, increase the threshold value as needed.
max_bytes_per_sec	The traffic limit on the raw data sent by Logtail. The value is of the int type and measured in bytes per second.	For example, the value 2097152 indicates that the data transfer rate of Logtail is restricted to 2 MB/s.
process_thread_count	The number of threads that Logtail uses to write data to log files.	The default value is 1. Typically, one thread allows a write speed of 24 MB/s in simple mode and 12 MB/s in full regex mode. Increase the threshold value only when necessary.

Parameter	Value	Description
send_request_concurrency	By default, Logtail sends data packets asynchronously. You can set a greater asynchronous concurrency value if the number of transactions per second (TPS) is large.	By default, four asynchronous concurrencies are available. You can calculate the quantity required based on the fact that each concurrency can provide 0.5 MB/s to 1 MB/s network throughput. The actual concurrency quantity varies with the network delay.
streamlog_open	Specifies whether to enable the syslog reception feature. The value is of the bool type.	False indicates that syslog reception is disabled and true indicates that syslog reception is enabled.
streamlog_pool_size_in_mb	The size of memory pool that the syslog uses to receive logs. The memory pool is used to cache syslog data. Unit: MB.	Logtail requests memory when it starts. Set the pool size based on the machine memory size and your actual needs.
streamlog_recv_size_each_call	The cache size that Logtail uses every time when calling the Linux socket recv operation. Unit: Bytes. Valid values: 1024 to 8192.	You can increase the value in the case of heavy syslog traffic.
streamlog_formats	The method of parsing received syslogs.	--
streamlog_tcp_addr	The binding address that Logtail uses to receive syslogs. Default value: 0.0.0.0.	For more information, see *Reference for collecting syslog data*.
streamlog_tcp_port	The TCP port that Logtail uses to receive syslogs.	Default value: 11111.

Parameter	Value	Description
buffer_file_num	**When a network exception occurs or the write quota is exceeded, Logtail writes the logs that are parsed in real time to a local file (in the installation directory) and then tries to resend the logs to Log Service after recovery. This parameter specifies the maximum number of cached files.**	**Default value: 25**
buffer_file_size	**The maximum number of bytes that a cached file allows.** `buffer_file_num * buffer_file_size` **indicates the maximum disk space available for cached files.**	**Default value: 20971520 Bytes (20 MB).**
buffer_file_path	**The directory that stores cached files. After modifying this parameter, you must manually move the files named** *logtail_buffer_file_** **in the old cache directory to the new cache directory so that Logtail can read the cached files and delete them after sending.**	**The default value is null, which indicates the cached files are stored in the Logtail installation directory** */usr/local/ilogtail*.

Parameter	Value	Description
bind_interface	The name of the NIC bound to the local machine, for example, eth1. This parameter is valid only for Logtail for Linux.	By default, the available NICs are bound automatically. If you specify this parameter, Logtail will forcibly use the specified NIC to upload logs.
check_point_filename	The full path for storing the checkpoint file. It is used to customize the storage path of checkpoint files of Logtail.	The default value is *`/tmp/logtail_check_point`*. If you use Docker, we recommend that you modify the storage path of the checkpoint file and mount the directory where the checkpoint file resides to the host. Otherwise, duplicate collection occurs due to missing checkpoint information when a container is released. For example, configure check_point_filename in Docker as `/data/logtail/check_point.dat`, and add `-v /data/docker1/logtail:/data/logtail` to Docker startup commands. Then, mount the *`/data/docker1/logtail`* directory of the host to the *`/data/logtail`* directory in Docker.

Note:

- The table only lists the important common startup parameters. If *`ilogtail_config.json`* includes parameters that are not listed in the table, the default values are used.
- You can add or modify the specified configuration parameters as needed. You do not need to add unnecessary configuration parameters to *`ilogtail_config.json`*.

Modify configurations

1. Configure *`ilogtail_config.json`* as needed.

   Confirm that the modified configurations are in the valid JSON format.

2. **Restart Logtail to apply the modified configurations.**

```
/etc/init.d/ilogtaild stop
/etc/init.d/ilogtaild start
/etc/init.d/ilogtaild status
```

18.7.3 Logtail machine group

18.7.3.1 Machine groups

Log Service manages all ECS instances whose logs need to be collected by Logtail in machine groups.

A machine group is a virtual group that contains multiple servers. If you have multiple servers and want to collect logs with the same Logtail configurations, you can add these servers to a machine group and apply the Logtail configuration to the machine group.

You can use the following methods to define a machine group:

- **IP address: Add the IP addresses of all servers to a machine group. Each server in the group can be identified by using its unique IP address.**
- **User-defined identifier: Customize an identifier for a machine group and use the same identifier for servers in the machine group.**

IP address-based machine groups

You can add multiple servers to a machine group by adding their IP addresses to the machine group. Then you can configure the Logtail clients on all the servers at the same time.

- **If you use ECS instances that are not bound with hostnames and the network types of these instances remain unchanged, you can use private IP addresses of these instances to define the machine group.**
- **In other cases, use the server IP address retrieved automatically by the Logtail client when you define a machine group. The IP address of each server is recorded in the IP address field of the `app_info.json` file on the server.**

Note:

`app_info.json` records the internal information of the Logtail client, which includes the server IP addresses automatically retrieved by the Logtail client. If you manually modify the IP address field of the file, the IP addresses obtained by the Logtail client remain unchanged.

A Logtail client automatically obtains a server IP address by using the following methods:

- If the IP address of a server has been bound with its hostname in the `/etc/hosts` file of a server, the Logtail client automatically retrieves the IP address.
- If the IP address of a server has not been bound with its hostname, the Logtail automatically retrieves the IP address of the first NIC on the server.

For more information about how to create an IP address-based machine group, see *Create an IP address-based machine group*.

Machine groups defined by user-defined identifiers

In addition to IP addresses, you can use user-defined identifiers to dynamically define machine groups. Multiple servers in a machine group can use one user-defined identifier to implement machine group automatic scaling. You only need to configure one user-defined identifier for new machines. Log Service then automatically identifies these machines and adds them to the machine group.

Typically, the system is composed of multiple modules. Each module can be scaled out separately, or multiple servers can be added to each module. By creating a machine group for each module, you can collect logs by module. Therefore, you need to create a user-defined identifier for each module, and set the machine group identifier for the servers of each module. For example, a typical website generally consists of an HTTP request processing module, a cache module, a logic processing module, and a storage module. Their user-defined identifiers can be defined as http_module, cache_module, logic_module, and store_module.

For more information about how to create a machine group with a user-defined identifier, see *Create a machine group with a user-defined identifier*.

18.7.3.2 Create an IP address-based machine group

IP address-based machine groups can be created with Log Service. After adding the IP addresses of servers retrieved by Logtail to an IP address-based machine group, you can use the same Logtail configuration to collect logs from the servers.

Prerequisites

- A project and a Logstore are created.
- One or more ECS instances are created.
- Logtail is installed on the servers.

Procedure

1. **Check the IP addresses of servers that are automatically retrieved by Logtail.**

   The IP addresses automatically retrieved by Logtail are recorded in the IP address field of the `app_info.json` **file.**

   On the server with Logtail installed, check the `app_info.json` **file. The file paths are as follows:**

   - **Linux:** `/usr/local/ilogtail/app_info.json`
   - **64-bit Windows:** `C:\Program Files (x86)\Alibaba\Logtail\app_info.json`
   - **32-bit Windows:** `C:\Program Files\Alibaba\Logtail\app_info.json`

2. *Log on to the Log Service console*.
3. **In the left-side navigation pane, choose LogHub - Real-time Collection > Logtail Machine Groups to go to the Machine Groups page of the project.**
4. **Click Create Machine Group in the upper-right corner.**

   Alternatively, after you create a collection configuration by using the data import wizard, click Create Machine Group on the Apply to Machine Group page.
5. **Create a machine group.**

   a) **Specify Machine Group Name.**

   The machine group name must be 3 to 128 characters in length and can contain only lowercase letters, digits, hyphens (-), and underscores (_). It must start and end with a lowercase letter or digit.

   > **Note:**
   > **After the machine group is created, you cannot modify its name. Proceed with caution.**

   b) **Set Machine Group Identifier to IP Address.**

   c) **Specify IP Address.**

   Enter the IP addresses of the servers obtained in Step 1.

   > **Note:**
   >
   > - **Make sure that you have obtained the IP addresses of servers based on** *Step 1*.

- If the machine group contains multiple servers, use line breaks to separate the IP addresses of the servers.

6. Optional: Set Machine Group Topic.
7. Click OK.

What's next

After the machine group is created, you can view the list of machine groups and the status of the machine group, manage configurations, and modify or delete the machine group.

18.7.3.3 Create a machine group with a user-defined identifier

Aside from IP addresses, you can also use user-defined identifiers to dynamically define machine groups.

User-defined identifiers have distinct advantages in the following scenarios:

- In a user-defined network environment such as a VPC, the IP address conflicts occur, resulting in failure of Logtail management. User-defined identifiers can help avoid such issues.
- Multiple machines can share the same tag to implement automatic scaling of machine groups. You only need to configure one user-defined identifier for new machines. Log Service then automatically identifies these machines and adds them to the machine group.

Procedure

1. Configure user-defined identifiers.

   Configure user-defined identifiers in the `/etc/ilogtail/user_defined_id` file.

   For example, set a user-defined machine identifier as follows:

   ```
   #cat /etc/ilogtail/user_defined_id
   ```

   Note:

   - You can configure multiple user-defined identifiers for a server and separate them with line breaks.

- If you want to use IP addresses to identify machines, delete the user_defined_id file. The new configuration takes effect in one minute. Run the following command to delete the user_defined_id file:

  ```
  rm -f /etc/ilogtail/user_defined_id
  ```

- If directories (*/etc/ilogtail/* or *C:\LogtailData*) or files (*/etc/ilogtail/user_defined_id* or *C:\LogtailData\user_defined_id*) do not exist, create them manually.

After you add, delete, or modify the user_defined_id file, the new configuration takes effect in one minute.

If you want to bring the new configuration into effect immediately, run the following commands to restart Logtail:

```
/etc/init.d/ilogtaild stop
/etc/init.d/ilogtaild start
```

2. Create a machine group.

   a) *Log on to the Log Service console*.

   b) Click the name of a project to go to the Logstores page.

   c) On the Machine Groups page, click Create Machine Group in the upper-right corner.

   d) Configure parameters for the machine group.

   - Machine Group Name: Specify the name of the machine group.

     The machine group name must be 3 to 128 characters in length and can contain only lowercase letters, digits, hyphens (-), and underscores (_). It must start and end with a lowercase letter or digit.

      **Note:**

After the machine group is created, you cannot modify its name. Proceed with caution.

- **Machine Group Identifier: Select User-defined Identifier.**
- **(Optional) Machine Group Topic: Enter a machine group topic.**
- **User Defined Identifier: Enter the user-defined identifier configured in step 1.**

e) **Click OK. To scale out machines, you just need to configure user-defined identifiers for new servers.**

3. **View machine group status.**

   On the Machine Groups page, click Machine Status to the right of the machine group to view the machines that use the same user-defined identifier and their heartbeats.

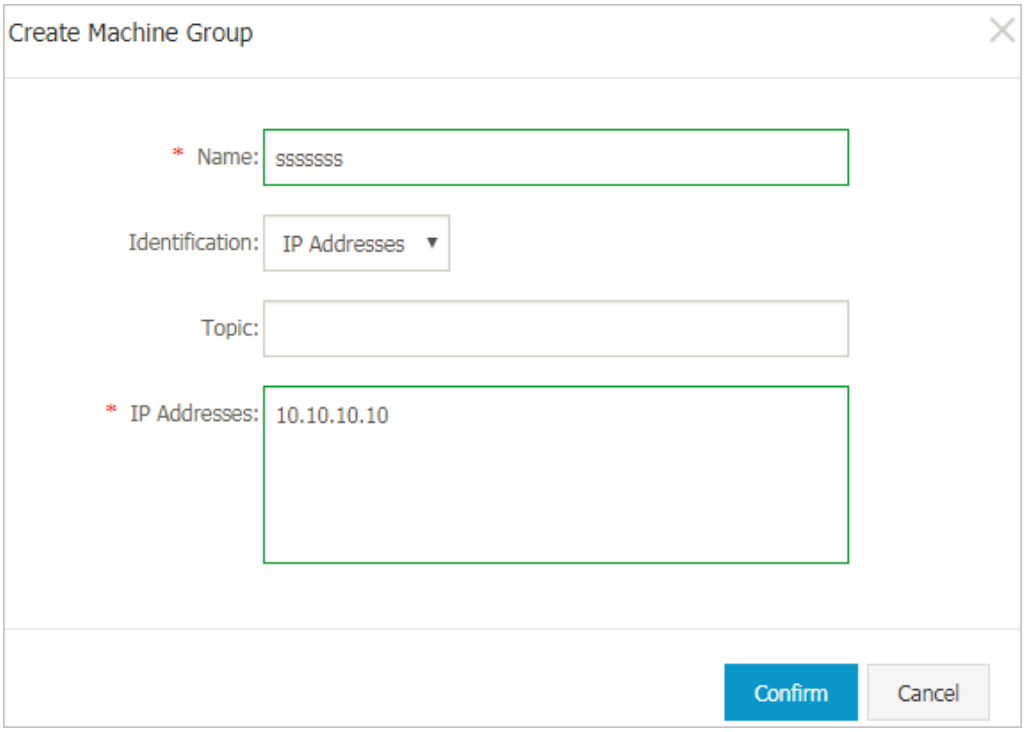
The system is comprised of multiple modules, each of which can contain multiple machines. For example, a typical website consists of a front-end HTTP request processing module, cache module, logic processing module, and storage module. Each module can be scaled out separately. When new machines are added, logs must be collected in real time.

1. **Create a user-defined identifier.**

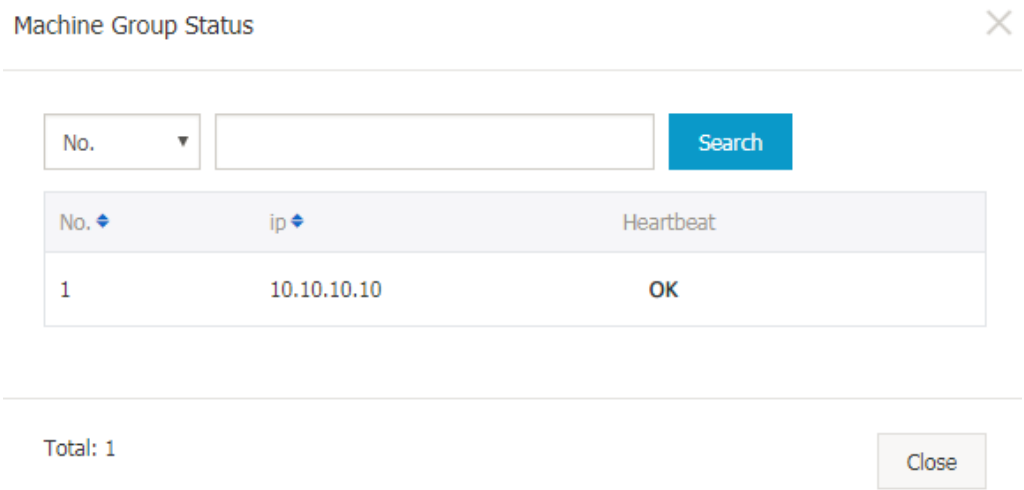
   After you install the Logtail client, enable user-defined identifier for the server. The user-defined identifiers of the modules in the preceding example can be defined as http_module, cache_module, logic_module, and store_module.

2. **Create a machine group.**

   When you create a machine group, enter the user-defined identifier of the machine group in the User-defined Identifier field. The following figure shows the http_module machine group.

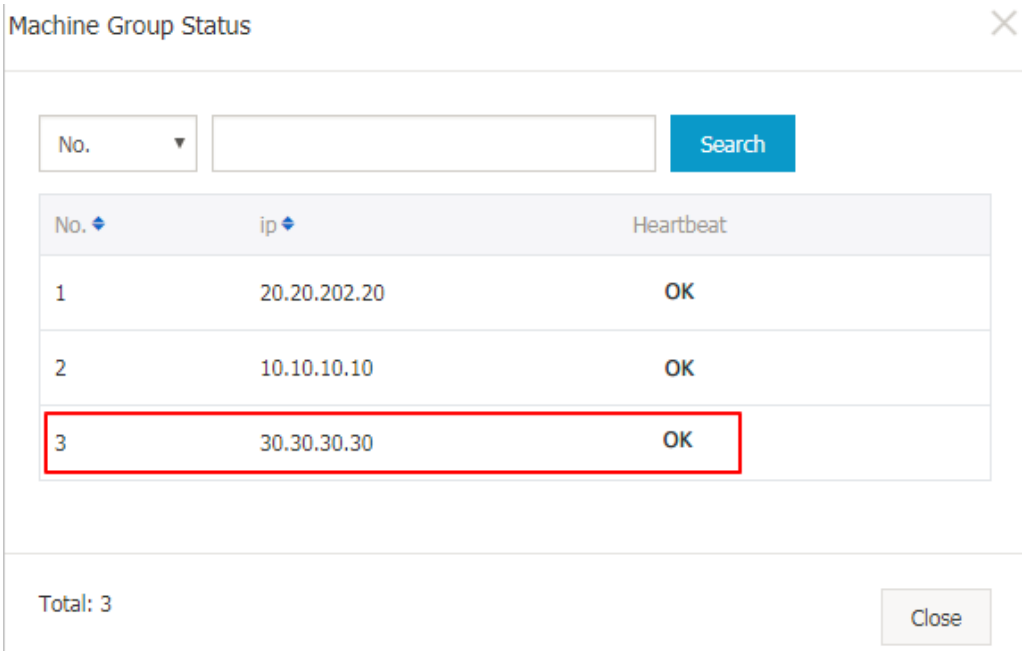


3. **Click Check Status for the machine group to view the list of machines that use the same user-defined identifier and their heartbeats.**



4. **If you add a machine with the IP address 10.1.1.3 to the front-end module, configure the user-defined identifier for the new server.**

After the identifier is configured, you can view the added server in the Machine Group Status dialog box.



18.7.3.4 View machine groups

On the Machine Groups page, you can view the machine groups created within the current project.

Prerequisites

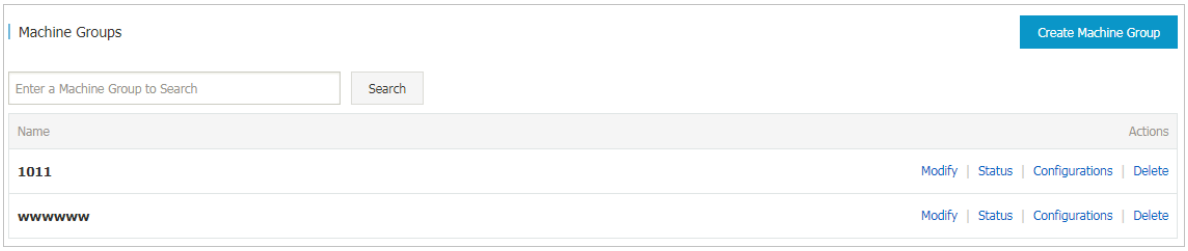
1. *Create a project*.
2. *Create a Logstore*.
3. *Create an IP address-based machine group*.

Procedure

1. *Log on to the Log Service console*.
2. **Click the target project name to go to the corresponding Logstores page.**

3. **In the left-side navigation pane, click Logtail Machine to go to the Machine Groups page.**

   All machine groups within the project can be viewed.



18.7.3.5 Modify a machine group

After creating a machine group, you can adjust the list of ECS instances in the machine group as needed.

Prerequisites

1. *Create a project*.
2. *Create a Logstore*.
3. *Create an IP address-based machine group*.

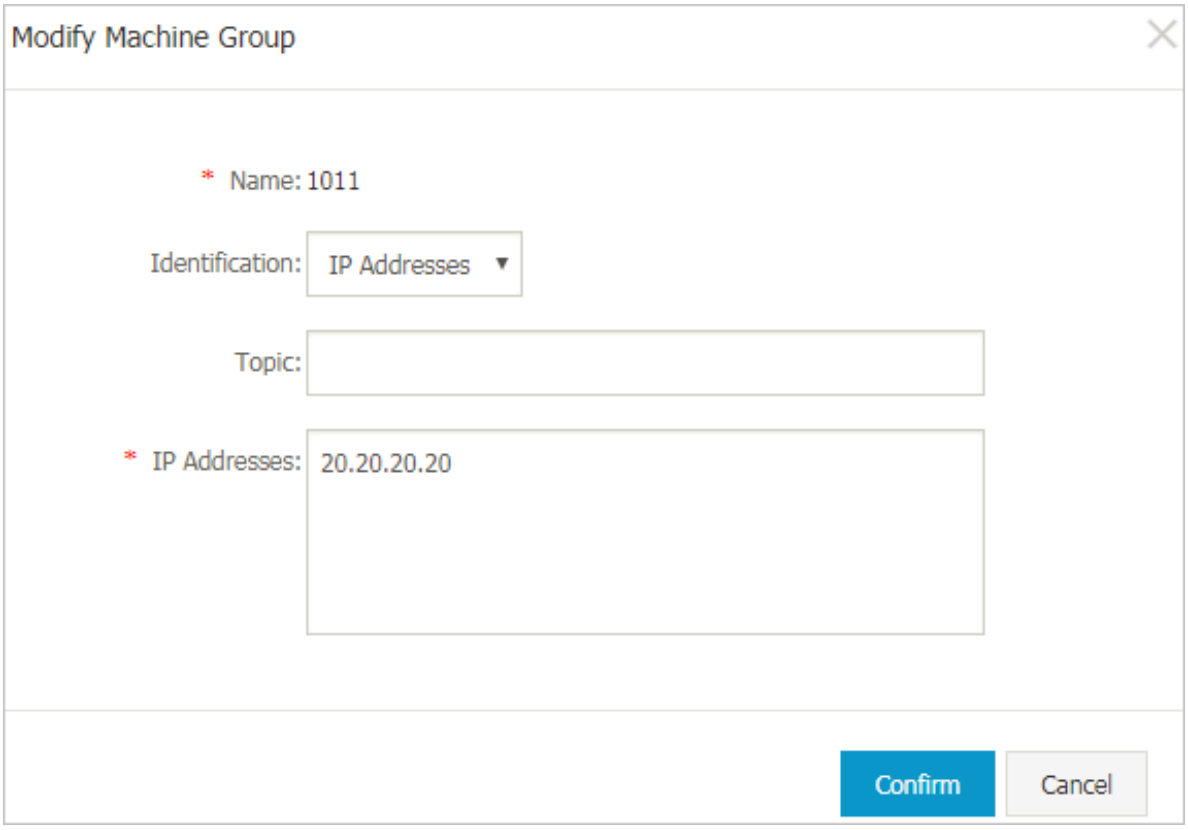
Modify a machine group

Procedure

1. *Log on to the Log Service console*.
2. **Click a project name to go to the Logstores page.**
3. **In the left-side navigation pane, click Logtail Machine to go to the Machine Groups page.**
4. **Select the machine group that you want to modify and click Modify.**

> **Note:**
> **The name of a machine group cannot be modified after the machine group is created.**

5. **Modify the machine group configuration, and then click Confirm.**



18.7.3.6 View machine group status

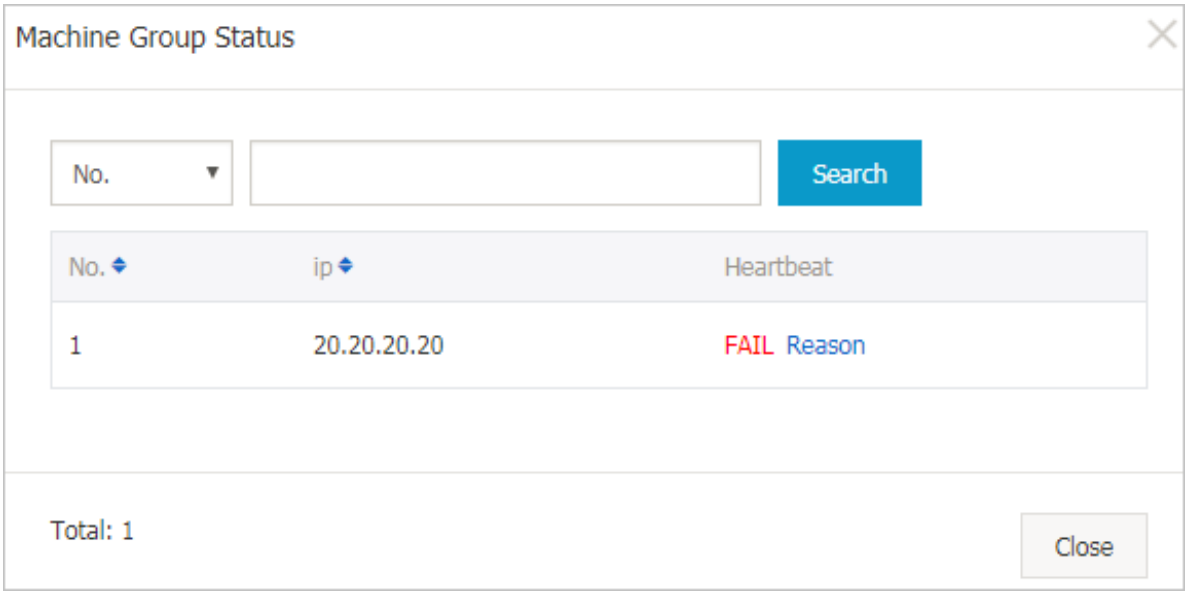
You can view the heartbeat information of the Logtail client to check whether the Logtail client is successfully installed on all ECS instances in a machine group.

Procedure

1. *Log on to the Log Service console*.
2. **Click a project name to go to the Logstores page.**
3. **In the left-side navigation pane, click Logtail Machine to go to the Machine Groups page.**
4. **Select a machine group and click Status.**

   If the Logtail client is successfully installed on all ECS instances, the heartbeat status of each ECS instance should be `OK`**. If the heartbeat status is** `FAIL`**, we**

recommend that you check the configuration first as prompted. If the issue cannot be resolved, submit a ticket for assistance.



18.7.3.7 Manage machine group configurations

Log Service uses machine groups to manage all ECS instances whose logs need to be collected by Logtail. You can access the Machine Groups page of a project through the Projects page of Log Service. Log Service allows you to create, modify, and delete machine groups, view the list and status of machine groups, manage configurations, and apply machine group IDs.

Prerequisites

1. *Create a project*.
2. *Create a Logstore*.
3. *Create an IP address-based machine group*.

Context

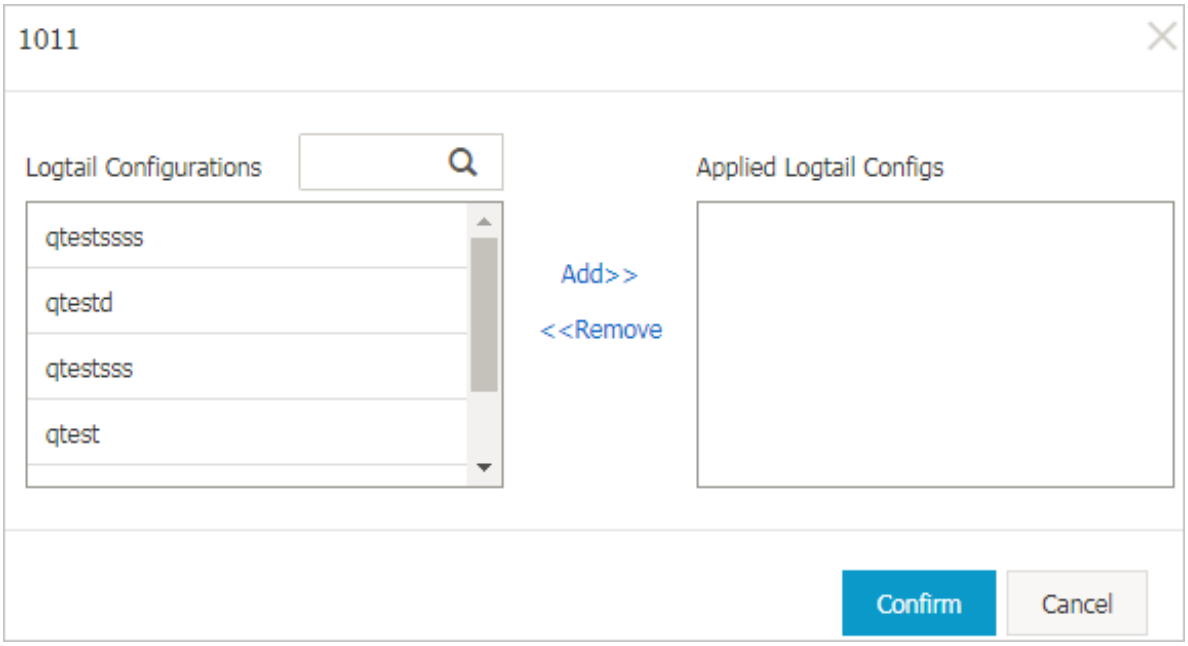
Log Service manages all ECS instances whose logs need to be collected through machine groups. Managing the collection configuration of the Logtail client is significant. For more information, see *Collect text files through Logtail* and *Collect syslog data through Logtail*. You can add or delete a Logtail configuration for a machine group to determine what logs are collected, how the logs are parsed, and to which Logstore the logs are sent by Logtail on each ECS instance.

Procedure

1. *Log on to the Log Service console*.

2. Click a project name to go to the Logstores page.
3. In the left-side navigation pane, click Logtail Machine to go to the Machine Groups page.
4. Select a machine group and click Configurations.
5. Select a Logtail configuration and click Add or Remove to modify the Logtail configurations applied to the machine group.

   After a Logtail configuration is added, the configuration is pushed to the Logtail client on each ECS instance in the machine group. When you remove a Logtail configuration, the Logtail configuration is removed from the Logtail client.



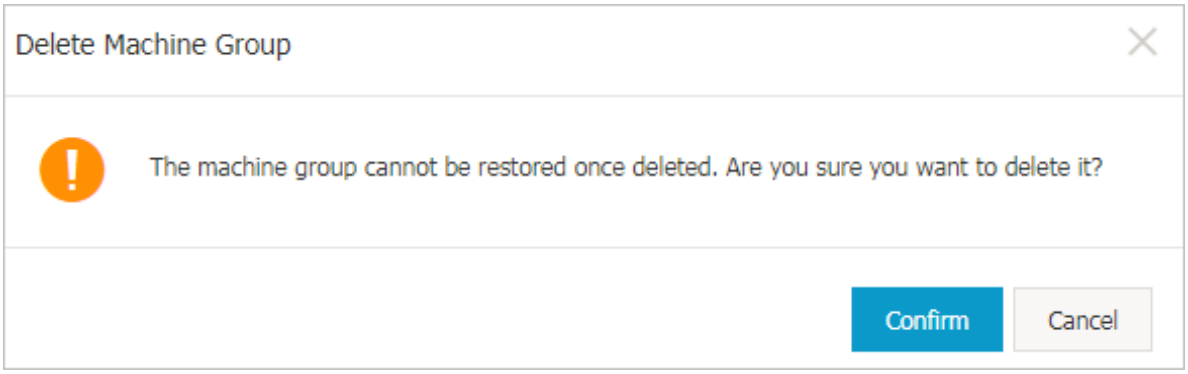
18.7.3.8 Delete a machine group

You can delete a machine group when you no longer need to collect logs from the machine group.

Procedure

1. *Log on to the Log Service console*.
2. Create a project and a Logstore. For more information, see *Create a project* and *Create a Logstore*.
3. Click a project name to enter the Logstores page. On the left-side navigation pane, click Logtail Machine. The Machine Groups page is displayed.
4. Select a machine group and click Delete.

5. In the dialog box that appears, click Confirm.



18.7.4 Data sources

18.7.4.1 Text logs

18.7.4.1.1 Collect text logs

The Logtail client can help you collect logs from ECS instances by using the Log Service console.

Context

After you create a Logstore, the system prompts you to access the data import wizard. In the dialog box that appears, click Data Import Wizard to create a Logtail configuration. Alternatively, you can go to the Logstores page and click the icon in the Data Import Wizard column to create a Logtail configuration.

Prerequisites

You must install Logtail before it can be used to collect logs. Apsara Stack Log Service allows you to install Logtail in Linux operating systems. For more information about the installation methods, see *Install Logtail in Linux*.

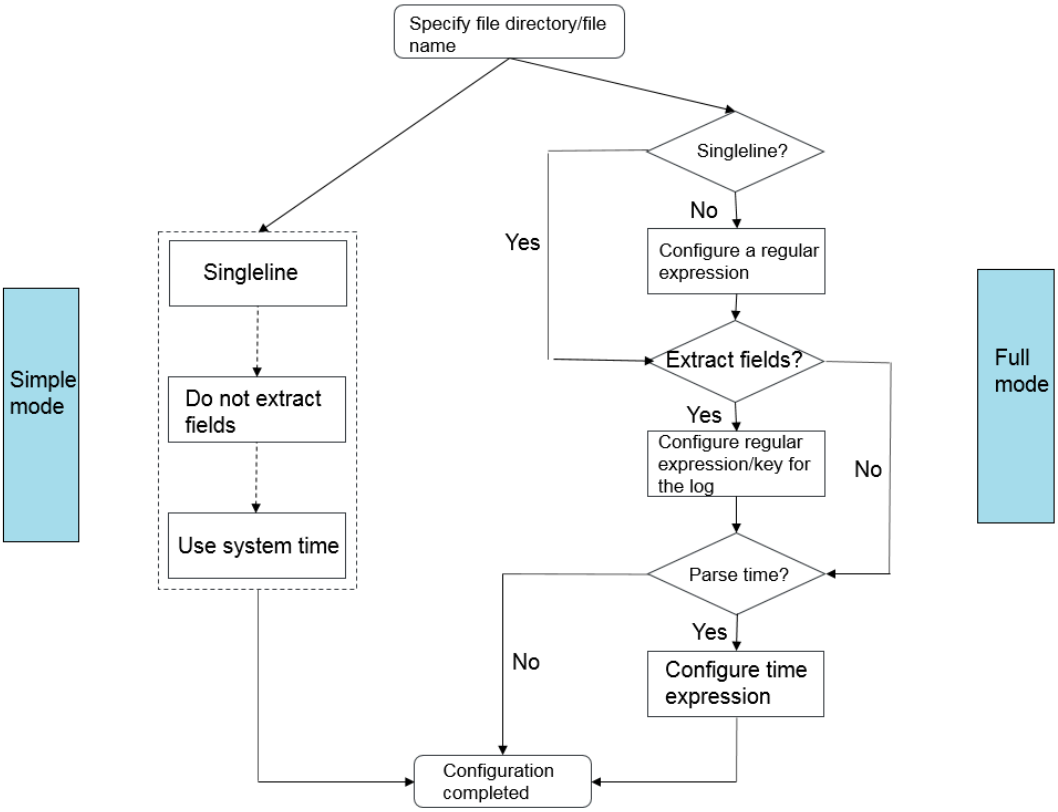
Limits

- **A log file can only be collected by a single configuration. To use multiple configurations to collect a log file, we recommend that you create a soft link. For example, you have a log file in the** `/home/log/nginx/log` **directory. To collect this file using two different configurations, use the original path for one configuration, and then run the** `ln -s /home/log/nginx/log /home/log/nginx/link_log` **command to create a soft link for the directory that can be used for the other configuration.**

- **For more information about operating systems that support the Logtail client, see** *Install Logtail in Linux*.

Log collection configuration

In the Log Service console, you can configure Logtail to collect text logs in different modes such as simple mode, delimiter mode, JSON mode, and full regex mode. The following figure shows the procedure of configuring Logtail to collect text logs in simple mode and full regex mode.



Procedure

1. *Log on to the Log Service console*.
2. **Create a project and a Logstore. For more information about how to create a project and a Logstore, see** *Create a project* **and** *Create a Logstore*.
3. **In the Log Service console, click the target project to go to the Logstores page.**
4. **Find the target Logstore and click the icon in the Data Import Wizard column corresponding to the Logstore.**
5. **Select a data type.**

   Select Text File under Custom Data and then click Next to go to the Configure Data Source page.

6. Configure a data source.

   a) Specify the configuration name.

   The configuration name must be 3 to 63 characters in length and can contain lowercase letters, digits, hyphens (-), and underscores (_). It must start and end with a lowercase letter or digit.

    **Note:**
   This name cannot be modified after the configuration is created.

   b) Specify the directory and name for the log file.

   The directory structure supports both the complete path mode and the wildcard mode.

    **Note:**

   - Only asterisks (`*`) and question marks (`?`) can function as wildcards.
   - A log file can only be collected by a single configuration.

   The specified name of the log file can be a complete match or a partial match that includes wildcards. For more information about file naming rules, see *Wildcard matching*.

   Log files are searched in multi-level directory matching mode. All files with compliant file names in the specified directory and corresponding subdirectories can be monitored.

   - Example: `/apsara/nuwa/ … /*.log` indicates files whose suffix is `.log` in the `/apsara/nuwa` directory, including recursive subdirectories.
   - Example: `/var/logs/app_* … /*.log*` indicates files whose file names contain `.log` located in any directories (and their subdirectories) within the `/var/logs` directory that conform to the `app_*` format.

   c) Specify the log collection mode.

   Log Service supports the following log parsing modes: NGINX Configuration, Simple Mode, Delimiter Mode, JSON Mode, and Full Regex Mode. This example

describes the collection mode settings based on Simple Mode and Full Regex Mode.

- **Simple Mode**

  Simple Mode refers to the single-line mode. A single line of log data is considered as a log. Different logs are separated within a log file by line breaks. The default regular expression is (.*) and log fields are not extracted . Logtail records the system time of the current server as the log generation time. For more advanced settings, select Full Regex Mode and then adjust the relevant settings.

  After selecting Simple Mode, you only need to specify the file directory and file name. Logtail collects logs line by line without extracting log fields. Logtail records the server system time at which a log is collected as the time of the log.

- **Full Regex Mode**

  Full Regex Mode allows you to set more personalized field extraction settings, such as collection of cross-line logs and field extraction.

  A. Specify Log Sample.

  The Log Service console can automatically generate a regular expression based on the specified sample log. Make sure that the specified log has actually been generated.

  B. Turn off Singleline.

  The single-line mode is enabled by default. In this mode, logs are separated by line breaks. To collect cross-line logs, such as Java program logs, you must turn off Singleline and configure Regular Expression.

  C. Configure Regular Expression.

  Two options are provided: Auto Generate and Manually Input Regular Expression. After entering a sample log, click Auto Generate. The Log Service console then generates a regular expression. If a regular

expression cannot be generated, you can switch to the manual mode, enter a regular expression, and then verify it.

D. Configure Extract Field.

To separately analyze and process fields within a log, you can use the Extract Field function to convert the specified field to a key-value pair and then send the key-value pair to Log Service. Therefore, you must specify a regular expression to parse the log content.

The Log Service console provides two methods to specify a regular expression for parsing log content. The first is to automatically generate a regular expression through simple interactions. You can select the fields to be extracted in the sample log by highlighting them and then click Generate RegEx. The Log Service console then automatically generates a regular expression.

Automatically generated regular expressions are convenient but may not be optimal. In this case, you can manually enter a regular expression. You can click Manually Input Regular Expression to switch to the manual input mode. After entering the regular expression, click Validate to verify that the regular expression can parse and extract the content of the sample log.

After a regular expression is automatically generated or manually specified, you must set a key for each extracted field.

d) Set Use System Time.

By default, Use System Time is turned on. If you turn off this switch, you must specify a field as the time field during field extraction and name this field `time`. After specifying the `time` field, you can click Auto Generate in Time Format to generate a method for parsing this field. For more information about log time formats, see *Configure a time format*.

e) Optional: Set Advanced Options.

Set Local Cache, Upload Raw Log, Topic Generation Mode, Log File Encoding, Maximum Directory Monitoring Depth, Timeout, and Filter Configuration

as needed. If you do not have any special requirements, you can retain the default configurations.

Parameter	Description
Local Cache	This parameter specifies whether to turn on Local Cache. When Log Service is unavailable, logs can be cached to a local directory and then uploaded after Log Service recovers. By default, up to 1 GB logs can be cached.
Upload Raw Log	This parameter specifies whether to upload raw logs. If you enable raw logs to be uploaded, the raw log content is uploaded as the __raw__ field with the parsed log content.
Topic Generation Mode	· Null - Do not generate topic: By default, this mode is selected. In this mode, the topic is set to an empty string and you can query logs without specifying the topic. · Machine Group Topic Attributes: This mode differentiates log data generated in different frontend servers. · File Path RegEx: When this mode is selected, you must configure Custom RegEx to extract a part of the file path as the topic. This mode differentiates log data generated by different users or instances.
Custom RegEx	If you select File Path RegEx, you must enter a custom regular expression.
Log File Encoding	· utf8: specifies UTF-8 encoding. · gbk: specifies GBK encoding.
Maximum Directory Monitoring Depth	This parameter specifies the maximum level of the directories that are monitored when logs are collected from the log source. The value ranges from 0 to 1000. The value 0 indicates that only the directory at the current level is monitored.

Parameter	Description
Timeout	A log file will time out if it is not updated within a specified period of time. The following Timeout options are available: · Never: All log files are continuously monitored and never time out. · 30 Minute Timeout: If a log file is not updated within 30 minutes, the system considers the file timed out and stops monitoring it.
Filter Configuration	Only logs that meet all filtering conditions are collected. For example, `Key:level Regex:WARNING|ERROR` indicates that only WARNING and ERROR logs are collected. You can also filter logs that do not conform to a condition. For example, `Key:level Regex:^(?!.*(INFO|DEBUG))` indicates that INFO and DEBUG logs are not collected. For more information about similar examples, see *regex-exclude-word* and *regex-exclude-pattern*.

f) After completing the settings, click Next.

7. Select the target machine group and click Apply to Machine Group.

   You must create a machine group before performing this operation. For more information about how to create a machine group, see *Create a machine group*.

   Note:

   · It takes up to 3 minutes for a Logtail configuration to be pushed and take effect.
   · After creating Logtail configurations, you can view the Logtail configuration list and modify or delete Logtail configurations.

18.7.4.1.2 Configure a time format

Each log in Log Service must contain a timestamp. When collecting logs from users' log files, Logtail must extract the timestamp string in each log and parse the string

as a timestamp. Therefore, you need to specify a timestamp format to facilitate parsing.

Logtail in Linux supports all time formats provided by the strftime function. Logtail can parse and use the timestamp strings that can be expressed in the log formats defined by the strftime function.

The timestamp strings of logs have diverse formats. To make configuration easier, the following table lists the common log time formats supported by Logtail:

Format	Description	Example
%a	The abbreviated week day name.	Fri
%A	The week day name.	Friday
%b	The abbreviated month name.	Jan
%B	The month name.	January
%d	The day of the month in numeric format. Valid values: 01 to 31.	07, 31
%h	The abbreviated month name. The format %h is equivalent to `%b`.	Jan
%H	The hour in the 24-hour format.	22
%I	The hour in the 12-hour format.	11
%m	The month in numeric format.	08
%M	The minute in numeric format. Valid values: 00 to 59.	59
%n	A line break.	Line break
%p	The local time in the a.m. or p.m. format.	AM/PM

Format	Description	Example
%r	The time in the 12-hour format. The format %r is equivalent to `%I:%M:%S %p`.	11:59:59 AM
%R	The time expressed in hours and minutes. The format %R is equivalent to `%H:%M`.	23:59
%S	The second in numeric format. Valid values: 00 to 59.	59
%t	A tab.	Tab
%y	The year in numeric format (two digits). Valid values: 00 to 99.	04, 98
%Y	The year in numeric format (four digits).	2004, 1998
%z	The time zone or the abbreviated time zone.	-07:00, +0800
%C	The century in numeric format. Valid values: 00 to 99.	16
%e	The day of the month in numeric format. Valid values: 1 to 31. A single digit is preceded by a space.	7, 31
%j	The day of the year. Valid values: 001 to 366.	365
%u	The week day in numeric format. The valid value ranges from 1 to 7. The value 1 represents Monday.	2
%U	The week of the year. Sunday is the first day of the week. Valid values: 00 to 53.	23

Format	Description	Example
%V	The week of the year. Monday is the first day of the week. If the first week of a month contains four or more days, this is considered as the first week. Otherwise, the next week is considered as the first week. Valid values: 01 to 53.	24
%w	The week day in numeric format. The valid value ranges from 0 to 6. The value 0 represents Sunday .	5
%W	The week of the year. Monday is the first day of the week. Valid values: 00 to 53.	23
%c	The standard date and time.	To specify more information such as long date and short date, you can use the supported formats defined in the preceding texts for more precise expression.
%x	The standard date.	To specify more information such as long date and short date, you can use the supported formats defined in the preceding texts for more precise expression.
%X	The standard time.	To specify more information such as long date and short date, you can use the supported formats defined in the preceding texts for more precise expression.

Format	Description	Example
%s	The UNIX timestamp.	1476187251

18.7.4.1.3 Generate a topic

A topic is a custom field used to mark a batch of logs. Logs in one Logstore can be grouped by log topics. You can specify a topic when writing or querying logs.

A log is the minimum data unit processed in Log Service. It is defined in a semi-structured data model. The specific data model consists of topic, time, content, and source. For more information, see *Log Service overview* **.**

A topic is a custom field used to mark a batch of logs. Logs in one Logstore can be grouped by log topics. You can specify a topic when writing or querying logs. For example, access logs are marked by sites, and platform users can use user IDs as the log topics and write them into the logs. In this way, you can choose to only view your own logs based on the log topic when querying logs. If you do not need to group logs in a Logstore, specify the same log topic for all of the logs. The default value of this field is an empty string, which is also a valid topic.

Note:
You cannot set a topic for syslog data.

You can set or change topics in the Log Service console.

Topic generation mode

You can set a topic when using Logtail to collect logs or using APIs or SDKs to upload logs. The following topic generation modes are supported in the console: Null - Do not generate topic, Machine Group Topic Attributes, and File Path RegEx.

- **Null - Do not generate topic**

  When you configure Logtail to collect text logs in the console, the default log topic generation mode is Null - Do not generate topic, that is, the topic is an empty string and you can query logs without entering a topic.
- **Machine Group Topic Attributes**

  The Machine Group Topic Attributes mode is used to differentiate log data generated on different servers. If log data of different servers is stored in the same file path and the same file, you can divide servers into different machine groups when you want to differentiate the log data of different servers by

topic. That is, set Group Topic differently for different machine groups when creating machine groups and set Topic Generation Mode to Machine Group Topic Attributes. Apply the previously created Logtail configuration to the machine groups to complete the configuration.

If Machine Group Topic Attributes is selected, Logtail uploads the topic attribute of the machine group to which the current machine belongs as the topic name to Log Service when reporting data. When you query logs by using LogSearch /Analytics, you need to specify a topic (namely the topic attribute of the target machine group) as the query condition.

- **File Path RegEx**

  The File Path RegEx mode is used to differentiate log data generated by users or instances. In some cases, the system stores logs in different directories for different users or instances but uses the same names for sub-directories and log files in these directories. As a result, Log Service cannot identify the user or instance for which the logs are generated when collecting log files. In these cases, you can set Topic Generation Mode to File Path RegEx, enter a regular expression of the file path, and set the topic to the instance name.

  When File Path RegEx is selected, Logtail uploads the instance name as the topic name to Log Service when reporting data. Different topics are generated according to your directory structure and configuration. You must specify the topic name as the instance name when querying logs by using LogSearch/ Analytics.

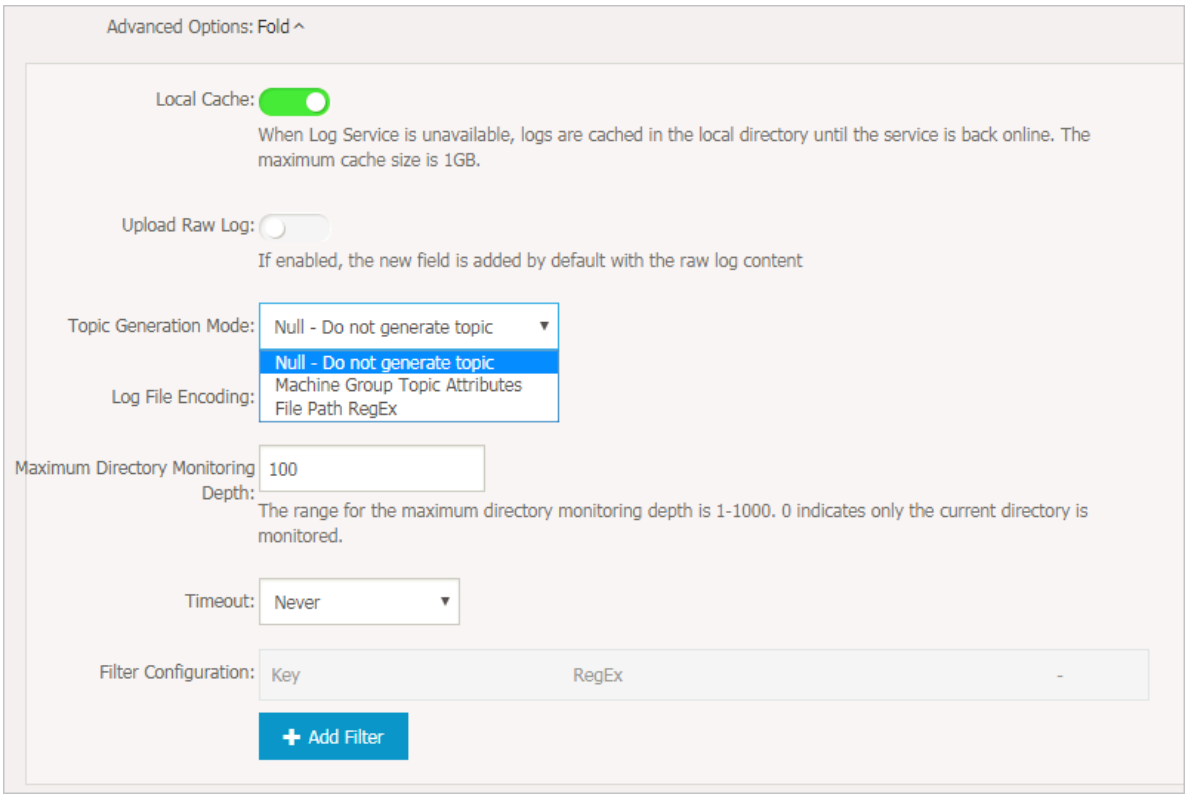
Set a log topic

Procedure

1. **Configure Logtail in the console according to** *Collect text logs*.

   To set the topic generation mode to Machine Group Topic Attributes, configure Machine Group Topic on the machine group creation and modification page.

2. Expand Advanced Options in the data import wizard and select a topic generation mode from the Topic Generation Mode drop-down list.



18.7.4.1.4 Import historical logs

Logtail collects only incremental logs by default. If you want to import historical logs, use the historical log importing feature of Logtail.

Prerequisites

- **The Logtail version must be `0.16.6` or later.**
- **Target historical logs must be in the configured collection range, and they have not been collected by Logtail.**
- **The last modification time of historical logs must be earlier than the Logtail configuration time.**
- **The maximum interval between generating and importing local events is 1 minute.**
- **Because loading local configurations is a special action, Logtail notifies you of this action by sending `LOAD_LOCAL_EVENT_ALARM` to your server.**

Context

Logtail collects files on an event-triggered basis. The system captures events by listening on the files or by checking the files for updates at intervals. Additional

ly, Logtail can load events from local files to trigger log collection. Historical file collection is a function implemented based on loading of local events.

Procedure

1. **Configure the collection.**

   Configure the collection and apply the configuration to the machine group. Make sure that the target logs are in the configured collection range. For more information about the collection configuration, see *Collect text logs*.

2. **Obtain a unique identifier for collection configuration.**

   Obtain a unique identifier for collection configuration from the local file `/usr/local/ilogtail/user_log_config.json`. The following content is used as an example:

   ```
   grep "##" /usr/local/ilogtail/user_log_config.json | awk '{print $1
   }'
                       "##1.0##log-config-test$multi"
                       "##1.0##log-config-test$ecs-test"
                       "##1.0##log-config-test$metric_system_test"
                       "##1.0##log-config-test$redis-status"
   ```

3. **Add local events.**

   Save local events to the JSON file `/usr/local/ilogtail/local_event.json` by using the following format:

   ```
   [
                           {
                           "config" : "${your_config_unique_id}",
                           "dir" : "${your_log_dir}",
                           "name" : "${your_log_file_name}"
                           },
                           {
                           ...
                           }
                           ...
                           ]
   ```

   - **Parameters**

Parameter	Description	Example
`config`	The unique identifier that is obtained in step 2.	##1.0##log-config-test$ecs-test

Parameter	Description	Example
dir	**The folder where logs are stored.** **Note:** **The folder cannot end with a slash (/).**	/data/logs
name	**The log name.**	access.log. 2018-08-08

Note:
To prevent Logtail from loading invalid JSON files, save local event configurations to a temporary file, edit the configurations in the temporary file, and copy the configuration content to the /usr/local/ilogtail/local_event.json file.

- **Sample configuration**

```
$ cat /usr/local/ilogtail/local_event.json
                                    [
                                    {
                                    "config": "##1.0##log-config-test$
ecs-test",
                                    "dir": "/data/log/",
                                    "name": "access.log. 2017-08-08"
                                    },
                                    {
                                    "config": "##1.0##log-config-test$
ecs-test",
                                    "dir": "/tmp",
                                    "name": "access.log. 2017-08-09"
                                    }
                                    ]
```

- **Check whether Logtail has loaded the configuration**

  After you save the `local_event.json` file, Logtail loads this local configuration file to the memory within 1 minute, and clears the content in `local_event.json`.

  You can check whether Logtail has read local events by using the following methods:

  - **Check whether the content in `local_event.json` has been cleared. If cleared, Logtail has read the local events.**
  - **Check whether the `/usr/local/ilogtail/ilogtail.LOG` file includes the keywords `process local event`. If the content in `local_event.json` has been cleared, but these keywords cannot be found, the local configuration file may be invalid and have been filtered out.**
- **Determine why Logtail has loaded the configuration but still cannot collect logs**

  This issue may be caused by the following reasons:

  - **The configuration is invalid.**
  - **The local `config` file does not exist.**
  - **The log file does not exist in the path specified in the collection configuration.**
  - **The logs have already been collected by Logtail.**
- **How to collect data that has already been collected**

  To collect data that has already been collected, follow these steps:

  1. **Run the `/etc/init.d/ilogtaild stop` command to stop Logtail.**
  2. **Find the path of the log file in the `/tmp/logtail_check_point` file.**
  3. **Delete the checkpoint (JSON object) of this log file and save the modification.**
  4. **Add the local event by following step 3 in the preceding section.**
  5. **Run the `/etc/init.d/ilogtaild start` command to start Logtail.**

18.7.4.2 Collect syslog logs

Logtail supports local configuration of TCP ports to receive syslog logs transferred through TCP by syslog agents. Logtail parses the received data and transfers it to LogHub.

Prerequisites

You must install Logtail before it can be used to collect logs. Apsara Stack Log Service allows you to install Logtail in Linux operating systems. For more information about the installation methods, see *Install Logtail in Linux*.

Step 1 Set Logtail syslog configuration parameters

1. *Log on to the Log Service console*.
2. **Create a project and a Logstore. For more information about how to create a project and a Logstore, see** *Project* **and** *Logstores*.
3. **In the Log Service console, click the target project to go to the corresponding Logstores page.**
4. **Find the target Logstore and click the icon in the Data Import Wizard column corresponding to the Logstore.**
5. **Select a data source type.**

   Select Syslog in Custom Data and click Next.
6. **Complete the syslog configuration and click Next.**

Parameter	Description
Configuration Name	The configuration name must be 3 to 63 characters in length, and can contain lowercase letters, digits, hyphens (-), and underscores (_). It must start and end with a lowercase letter or digit.  **Note:** This name cannot be modified after the configuration is created.
Tag Settings	For more information about tag settings, see *Reference for collecting syslog logs*.

Parameter	Description
Advanced Options	The following table lists the advanced configuration of syslog log collection.

Parameter	Description
Local Cache	This parameter specifies whether to turn on Local Cache. When Log Service is unavailable, logs can be cached to a local directory and then uploaded after Log Service recovers. By default, up to 1 GB logs can be cached.
Upload Raw Log	This parameter specifies whether to upload raw logs. If you enable raw logs to be uploaded, the raw log content is uploaded as the __raw__ field with the parsed log content .
Topic Generation Mode	· Null - Do not generate topic: By default, this mode is selected. In this mode, the topic is set to an empty string and you can query logs without specifying the topic. · Machine Group Topic Attributes: This mode differentiates log data generated in different frontend servers. · File Path RegEx: When this mode is selected, you must configure Custom RegEx to extract a part of the file path as the topic. This mode differentiates log data generated by different users or instances.
Custom RegEx	If you select File Path RegEx, you must enter a custom regular expression.
Log File Encoding	· utf8: specifies UTF-8 encoding. · gbk: specifies GBK encoding.
Maximum Directory Monitoring Depth	This parameter specifies the maximum level of the directories that are monitored when logs are collected from the log source. The value ranges from 0 to 1000. The value 0 indicates that only the directory at the current level is monitored.

Parameter	Description
Timeout	A log file will time out if it is not updated within a specified period of time. The following Timeout options are available: · Never: All log files are continuously monitored and never time out. · 30 Minute Timeout: If a log file is not updated within 30 minutes, the system considers the file timed out and stops monitoring it.
Filter Configuration	Only logs that meet all filtering conditions are collected. For example, `Key:level Regex:WARNING|ERROR` indicates that only WARNING and ERROR logs are collected. You can also filter logs that do not conform to a condition. For example, `Key:level Regex:^(?!.*(INFO|DEBUG))` indicates that INFO and DEBUG logs are not collected. For more information about similar examples, see *regex-exclude-word* and *regex-exclude-pattern*.

7. Apply the Logtail configuration to a machine group.

   Select a machine group and click Apply to Machine Group to apply the configuration to the selected machine group.

   You must create a machine group before performing this operation. For more information about how to create a machine group, see *Create a machine group*.

Step 2 Configure Logtail to enable syslog

Go to the Logtail installation directory `/usr/local/ilogtail/` on the server, find `ilogtail_config.json`, and modify the syslog-related settings as needed.

Procedure

1. **Check whether syslog is enabled.**

   true indicates that syslog is enabled, while false indicates that syslog is disabled.

   ```
   "streamlog_open" : true
   ```

2. **Configure the size of the syslog memory pool for storing received logs.**

   Logtail requests a specific size of memory when it is started. You can set the size of the syslog memory pool based on the memory size of the server and your business needs. Unit: MB.

   ```
   "streamlog_pool_size_in_mb" : 50
   ```

3. **Configure the buffer size.**

   You must specify the size of the buffer that Logtail uses when calling the socket io rcv interface. Unit: Byte.

   ```
   "streamlog_rcv_size_each_call" : 1024
   ```

4. **Configure the syslog log format.**

   ```
   "streamlog_formats":[]
   ```

5. **Configure the TCP port.**

   Configure the TCP port used by Logtail to receive syslog logs. By default, port 11111 is used.

   ```
   "streamlog_tcp_port" : 11111
   ```

6. **Restart Logtail.**

   To restart Logtail, run the following commands to stop the Logtail client and then start it again.

   ```
   sudo /etc/init.d/ilogtaild stop
   sudo /etc/init.d/ilogtaild start
   ```

7. **Install rsyslog.**

   For more information about installing rsyslog, visit the following websites:

   - *Install rsyslog on Ubuntu*
   - *Install rsyslog on Debian*
   - *Install rsyslog on RHEL or CENTOS*

8. **Modify the settings.**

   Navigate to `/etc/rsyslog.conf` and modify the settings in the file as needed.

   Example:

```
$WorkDirectory /var/spool/rsyslog # where to place spool files
$ActionQueueFileName fwdRule1     # unique name prefix for spool
files
$ActionQueueMaxDiskSpace 1g       # 1gb space limit (use as much as
possible)
$ActionQueueSaveOnShutdown on     # save messages to disk on
shutdown
$ActionQueueType LinkedList       # run asynchronously
$ActionResumeRetryCount -1        # infinite retries if host is down
# Defines the fields of log data
$template ALI_LOG_FMT,"0.1 sys_tag %timegenerated:::date-unixtimest
amp% %fromhost-ip% %hostname% %pri-text% %protocol-version% %app-
name% %procid% %msgid% %msg:::drop-last-lf%\n"
*. *      @@10.101.166.173:11111;ALI_LOG_FMT
```

   Note:
   In the `ALI_LOG_FMT` template, the value of the second field is `sys_tag`. This value must be the same as the one entered in step 1. This configuration indicates that all syslog logs in the `\*.\*` format received by the local server is formatted based on the `ALI_LOG_FMT` template and transferred to 10.101.166.173:11111 through TCP. The server with the address 10.101.166.173 must belong to the machine group selected in Step 1, and you have configured the server by following Step 2.

9. **Start rsyslog.**

```
sudo /etc/init.d/rsyslog restart
```

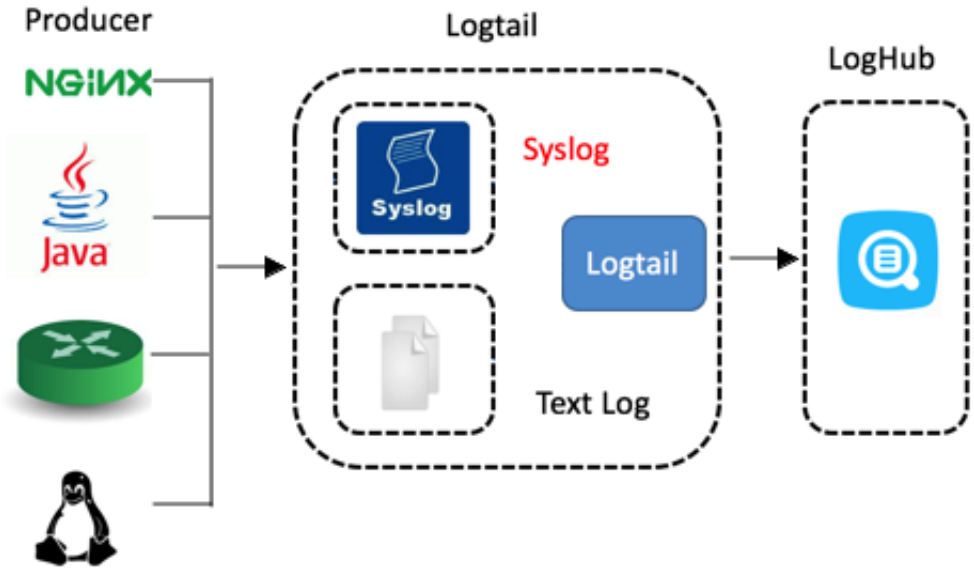
   Before starting rsyslog, check whether another syslog agent, such as syslogd , sysklogd, and syslog-ng, is installed on the server. If another syslog agent is installed, disable it.

   After the preceding steps, syslog logs on the server can be collected and sent to Log Service.

18.7.4.3 Reference for collecting syslog logs

This topic provides the reference for collecting syslog logs through Logtail.

Logtail supports collection of syslog and text logs, as shown in the following figure.



Logtail collects syslog logs based on TCP. For more information about how to configure Logtail to collect syslog logs, see *Collect syslog logs through Logtail*.

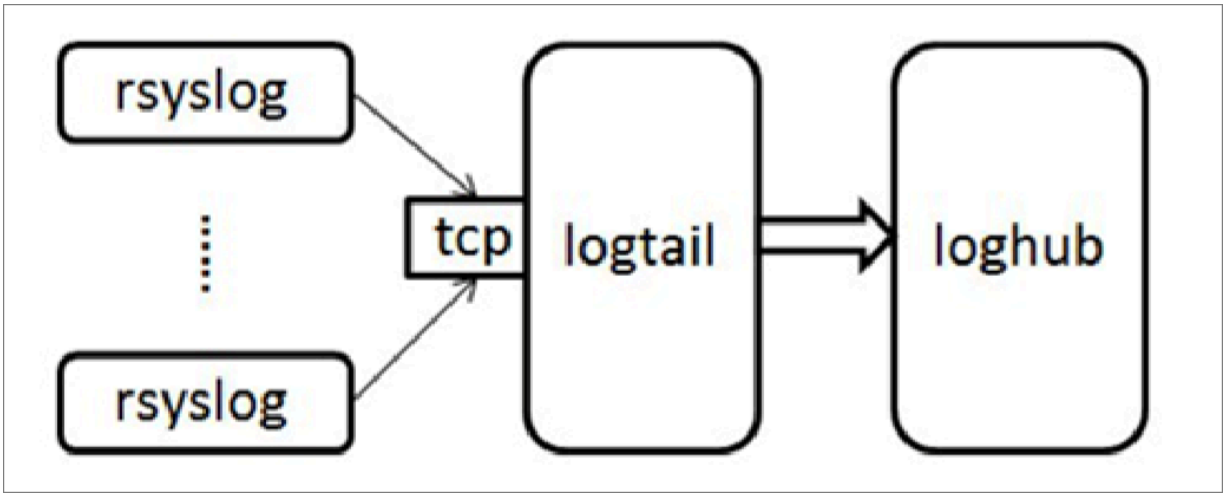
Advantages of syslog

For more information about syslog concepts, see *syslog*.

Compared with text logs, syslog logs are collected and directly transferred to LogHub without being stored into disks. This provides enhanced confidentiality and removes the need for data parsing.

How it works

Logtail supports local configuration of TCP ports to receive syslog logs transferre d by syslog agents. The following figure shows the relationship between Logtail, syslog, and LogHub. Logtail enables TCP ports to receive syslog logs transferred through TCP by rsyslog or other syslog agents. Logtail parses the received data and transfers it to LogHub.



Syslog log format

Logtail receives streaming data by using TCP ports. To parse individual logs from the streaming data, make sure that the log format meets the following requiremen ts:

- **Logs are separated by line breaks (`\n`). Line breaks cannot appear within any of the logs.**
- **Only the message body of a log can contain spaces. Other fields cannot contain spaces.**

Syslog logs are in the following format:

```
$version $tag $unixtimestamp $ip [$user-defined-field-1 $user-defined-field-2 $user-defined-field-n] $msg\n"
```

The following table describes fields in a syslog log.

Log field	Description
version	The version of the log format. Logtail uses the version to parse user-defined fields.
tag	The data tag used to locate the project or Logstore. The value cannot contain spaces or line breaks.
unixtimestamp	The timestamp of the log.
ip	The IP address of the server corresponding to the log. If the field value is 127.0.0.1, it is replaced with the IP address of the peer end of the TCP socket when the log is sent to Log Service.

Log field	Description
user-defined-field	Optional. The user-defined field. Zero or multiple user-defined fields can be set. The fields cannot contain spaces or line breaks.
msg	The message body of the log, which cannot contain line breaks. The `\n` that follows this field is the line break.

The following sample log meets the format requirements:

```
2.1 streamlog_tag 1455776661 10.101.166.127 ERROR com.alibaba.
streamlog.App.main(App.java:17) connection refused, retry
```

In addition to syslog, other log tools that meet the following requirements can also interconnect with Logtail:

- **Can format logs. The formatted logs must meet the format requirements.**
- **Can append logs to the remote end through TCP.**

Rules for Logtail to parse syslog logs

Logtail needs additional configurations for parsing syslog logs. Example:

```
"streamlog_formats":
[
    {"version": "2.1", "fields": ["level", "method"]},
    {"version": "2.2", "fields": []},
    {"version": "2.3", "fields": ["pri-text", "app-name", "syslogtag
"]}
]
```

Apply the following configuration: Logtail identifies the format of the correspond ing user-defined field in streamlog_formats based on the version field. The preceding sample log with the version field set to 2.1 contains two user-defined fields: level and method. Therefore, this log is parsed into the following format:

```
{
    "source": "10.101.166.127",
    "time": 1455776661,
    "level": "ERROR",
    "method": "com.alibaba.streamlog.App.main(App.java:17)",
    "msg": "connection refused, retry"
}
```

The version field is used to parse user-defined fields, and the tag field is used to search for the project or Logstore where the data is sent. These two fields are not included in the logs sent to Log Service. In addition, Logtail predefines some log

formats where the value of the version field starts with "0." or "1.", such as 0.1 and 1.1. Therefore, values of user-defined version fields cannot start with "0." or "1.".

Common log tools that can interconnect with Logtail

- **Log4j**
  - **Introduce the Log4j library.**

```
<dependency>
      <groupId>org.apache.logging.log4j</groupId>
      <artifactId>log4j-api</artifactId>
      <version>2.5</version>
  </dependency>
  <dependency>
      <groupId>org.apache.logging.log4j</groupId>
      <artifactId>log4j-core</artifactId>
      <version>2.5</version>
  </dependency>
```

  - **Introduce the Log4j configuration file** `log4j_aliyun.xml`.

```
 <? xml version="1.0" encoding="UTF-8"? >
  <configuration status="OFF">
      <appenders>
          <Socket name="StreamLog" protocol="TCP" host="10.101.166
.173" port="11111">
              <PatternLayout pattern="%X{version} %X{tag} %d{UNIX}
 %X{ip} %-5p %l %enc{%m}%n" />
          </Socket>
      </appenders>
      <loggers>
          <root level="trace">
              <appender-ref ref="StreamLog" />
          </root>
      </loggers>
  </configuration>
```

    In the preceding profile, 10.101.166.173:11111 is the IP address of the server where Logtail is located.

  - **Set ThreadContext in programs.**

```
package com.alibaba.streamlog;
  import org.apache.logging.log4j.LogManager;
  import org.apache.logging.log4j.Logger;
  import org.apache.logging.log4j.ThreadContext;
  public class App
  {
      private static Logger logger = LogManager.getLogger(App.
class);
      public static void main( String[] args ) throws Interrupte
dException
      {
          ThreadContext.put("version", "2.1");
          ThreadContext.put("tag", "streamlog_tag");
          ThreadContext.put("ip", "127.0.0.1");
          while(true)
          {
```

```
                logger.error("hello world");
                Thread.sleep(1000);
            }
            //ThreadContext.clearAll();
        }
    }
```

- **Tengine**

  Tengine can interconnect with ilogtail through syslog.

  Tengine uses ngx_http_log_module for recording logs to the local syslog agent. The local syslog agent forwards the logs to rsyslog.

  For more information about how to configure syslog in Tengine, see *Configure syslog in Tengine*.

  Example:

  Send info-level access logs of the user type to unix-dgram (/dev/log) of the local server and set the application tag to nginx.

  ```
  access_log syslog:user:info:/var/log/nginx.sock:nginx
  ```

  Rsyslog configuration:

  ```
  module(load="imuxsock") # needs to be done just once
  input(type="imuxsock" Socket="/var/log/nginx.sock" CreatePath="on")
  $template ALI_LOG_FMT,"2.3 streamlog_tag %timegenerated:::date-
  unixtimestamp% %fromhost-ip% %pri-text% %app-name% %syslogtag% %msg
  :::drop-last-lf%\n"
  if $syslogtag == 'nginx' then @@10.101.166.173:11111;ALI_LOG_FMT
  ```

- **NGINX**

  The following example describes the collection of NGINX access logs.

  Access log configuration:

  ```
  access_log syslog:server=unix:/var/log/nginx.sock,nohostname,tag=
  nginx;
  ```

  Rsyslog configuration:

  ```
  module(load="imuxsock") # needs to be done just once
  input(type="imuxsock" Socket="/var/log/nginx.sock" CreatePath="on")
  $template ALI_LOG_FMT,"2.3 streamlog_tag %timegenerated:::date-
  unixtimestamp% %fromhost-ip% %pri-text% %app-name% %syslogtag% %msg
  :::drop-last-lf%\n"
  if $syslogtag == 'nginx' then @@10.101.166.173:11111;ALI_LOG_FMT
  ```

  For more information, see *http://nginx.org/en/docs/syslog.html*.

- **Python syslog**

  Example:

```
import logging
import logging.handlers
logger = logging.getLogger('myLogger')
logger.setLevel(logging.INFO)
#add handler to the logger using unix domain socket '/dev/log'
handler = logging.handlers.SysLogHandler('/dev/log')
#add formatter to the handler
formatter = logging.Formatter('Python: { "loggerName":"%(name)s",
 "asciTime":"%(asctime)s", "pathName":"%(pathname)s", "logRecordC
reationTime":"%(created)f", "functionName":"%(funcName)s", "levelNo
":"%(levelno)s", "lineNo":"%(lineno)d", "time":"%(msecs)d", "
levelName":"%(levelname)s", "message":"%(message)s"}')
handler.formatter = formatter
logger.addHandler(handler)
logger.info("Test Message")
```

18.7.5 Troubleshooting

18.7.5.1 Query the local log collection status

You can view the Logtail health status and log collection progress. This helps you to check log collection issues and customize status monitoring for log collection.

Instructions

After a Logtail client that supports the status query function is installed, you can query local log collection status by entering commands on the client. For more information about how to install Logtail, see *Install Logtail in Linux*.

Enter the `/etc/init.d/ilogtaild -h` **command on the client to check whether the client supports the query of local log collection status. If the command output contains** `logtail insight, version`**, this function is supported on the Logtail client.**

```
/etc/init.d/ilogtaild -h
Usage: ./ilogtaild { start | stop (graceful, flush data and save
checkpoints) | force-stop | status | -h for help}$
logtail insight, version : 0.1.0
commond list :
       status all [index]
             get logtail running status
       status active [--logstore | --logfile] index [project] [
logstore]
             list all active logstore | logfile. if use --logfile,
please add project and logstore. default --logstore
       status logstore [--format=line | json] index project logstore
             get logstore status with line or json style. default --
format=line
       status logfile [--format=line | json] index project logstore
fileFullPath
```

```
            get log file status with line or json style. default --
format=line
      status history beginIndex endIndex  project logstore [
fileFullPath]
            query logstore | logfile history status.
index :   from 1 to 60. in all, it means last $(index) minutes; in
active/logstore/logfile/history, it means last $(index)*10 minutes
```

Logtail supports multiple query commands. The following table describes the query commands and the command-related information containing command functions, time intervals to query, and time windows for result statistics.

Command	Function	Time interval to query	Time window for statistics
all	Query the running status of Logtail.	Last 60 minutes	1 minute
active	Query Logstores or log files that are currently active (that is, with data collected).	Last 600 minutes	10 minutes
logstore	Query the collection status of a Logstore.	Last 600 minutes	10 minutes
logfile	Query the collection status of a log file.	Last 600 minutes	10 minutes
history	Query the collection status of a Logstore or log file over a period of time.	Last 600 minutes	10 minutes

Note:

- **The `index` parameter in the command indicates the index value of the time window, which is counted from the current time. The valid index value ranges from 1 to 60. If the time window for statistics is 1 minute, windows in the last `(index, index-1]` minutes are queried. If the time window for statistics is 10 minutes, windows in the last `(10 × index, 10 × (index - 1)]` minutes are queried.**

- **All query commands belong to status subcommands, so the main command is status.**

all

Command format

```
/etc/init.d/ilogtaild status all [ index ]
```

Note:
The all command is used to view the running status of Logtail. The index parameter is optional. If no value is entered, the default value 1 is used.

Example

```
/etc/init.d/ilogtaild status all 1
ok
/etc/init.d/ilogtaild status all 10
busy
```

Output description

Item	Description	Priority	Solution
ok	The current status is normal.	N/A	No action is required.
busy	The current collection speed is high, and Logtail is running properly.	N/A	No action is required.
many_log_files	A large number of log files are being collected.	Low	Check whether the configuration contains files that do not need to be collected.

Item	Description	Priority	Solution
process_block	Current log parsing is blocked.	Low	Check whether logs are generated at an excessively high speed. If this issue persists, *Configure startup parameters* as needed to modify the upper limit of CPU usage or limits on concurrent inbound traffic to Log Service.
send_block	Current sending is blocked.	Relatively high	Check whether logs are generated at an excessively high speed and whether the network status is normal. If this issue persists, *Configure startup parameters* as needed to modify the upper limit of CPU usage or limits on concurrent inbound traffic to Log Service.

active command

Command format

```
/etc/init.d/ilogtaild status active [--logstore] index
 /etc/init.d/ilogtaild status active --logfile index project-name
logstore-name
```

Note:

- The `active [--logstore] index` command is used to query Logstores that are currently active. The `--logstore` parameter can be omitted without changing the meaning of the command.

- **The `active --logfile index project-name logstore-name` command is used to query all active log files in a Logstore for a project.**
- **The active command is used to query active log files level by level. We recommend that you first locate the currently active Logstore and then query active log files in this Logstore.**

Example

```
/etc/init.d/ilogtaild status active 1
sls-zc-test : release-test
sls-zc-test : release-test-ant-rpc-3
sls-zc-test : release-test-same-regex-3

/etc/init.d/ilogtaild status active --logfile 1 sls-zc-test release-
test
/disk2/test/normal/access.log
```

Output description

- **After the `active --logstore index` command is executed, all currently active Logstores are returned in the format of `project-name : logstore-name`. After the `active --logfile index project-name logstore-name` command is executed, the complete paths of active log files are returned.**
- **A Logstore or log file with no log collection activity in the current query window does not appear in the output.**

logstore command

Command format

```
/etc/init.d/ilogtaild status logstore [--format={line|json}] index
project-name logstore-name
```

Note:

- **The logstore command is used to query the collection status of the specified project and Logstore in LINE or JSON format.**
- **If the `--format=` parameter is not configured, `--format=line` is selected by default. The output is displayed in LINE format.** *Note*: **The `--format` parameter must be placed after `logstore`.**
- **If this Logstore does not exist or has no log collection activity in the current query window, you get an empty output in LINE format or a `null` value in JSON format.**

Example

```
/etc/init.d/ilogtaild status logstore 1 sls-zc-test release-test-same
time_begin_readable : 17-08-29 10:56:11
time_end_readable : 17-08-29 11:06:11
time_begin : 1503975371
time_end : 1503975971
project : sls-zc-test
logstore : release-test-same
status : ok
config : ##1.0##sls-zc-test$same
read_bytes : 65033430
parse_success_lines : 230615
parse_fail_lines : 0
last_read_time : 1503975970
read_count : 687
avg_delay_bytes : 0
max_unsend_time : 0
min_unsend_time : 0
max_send_success_time : 1503975968
send_queue_size : 0
send_network_error_count : 0
send_network_quota_count : 0
send_network_discard_count : 0
send_success_count : 302
send_block_flag : false
sender_valid_flag : true
/etc/init.d/ilogtaild status logstore --format=json 1 sls-zc-test
release-test-same
{
   "avg_delay_bytes" : 0,
   "config" : "##1.0##sls-zc-test$same",
   "last_read_time" : 1503975970,
   "logstore" : "release-test-same",
   "max_send_success_time" : 1503975968,
   "max_unsend_time" : 0,
   "min_unsend_time" : 0,
   "parse_fail_lines" : 0,
   "parse_success_lines" : 230615,
   "project" : "sls-zc-test",
   "read_bytes" : 65033430,
   "read_count" : 687,
   "send_block_flag" : false,
   "send_network_discard_count" : 0,
   "send_network_error_count" : 0,
   "send_network_quota_count" : 0,
   "send_queue_size" : 0,
   "send_success_count" : 302,
   "sender_valid_flag" : true,
   "status" : "ok",
   "time_begin" : 1503975371,
   "time_begin_readable" : "17-08-29 10:56:11",
   "time_end" : 1503975971,
   "time_end_readable" : "17-08-29 11:06:11"
}
```

Output description

Keyword	Description	Unit
status	The overall status of this Logstore. For specific status, descriptions, and processing methods, see the following Logstore status table.	N/A
time_begin_readable	The start time of reading.	N/A
time_end_readable	The end time of reading.	N/A
time_begin	The start time of statistics collection.	UNIX timestamp, in seconds
time_end	The end time of statistics collection.	UNIX timestamp, in seconds
project	The project name.	N/A
logstore	The Logstore name.	N/A
config	The collection configuration name. It is a globally unique name which is made up of `##1.0##`, project, `$`, and config.	N/A
read_bytes	The number of logs read in the window.	Bytes
parse_success_lines	The number of successfully parsed log lines in the window.	Line
parse_fail_lines	The number of log lines that failed to be parsed in the window.	Line
last_read_time	The last reading time in the window.	UNIX timestamp, in seconds
read_count	The number of times logs are read in the window.	Number of times
avg_delay_bytes	The average of the differences between the current offset and the file size each time logs are read in the window.	Bytes

Keyword	Description	Unit
max_unsend_time	The maximum time that unsent data packets are in the send queue when the window closes. The value is 0 when the queue is empty.	UNIX timestamp, in seconds
min_unsend_time	The minimum time that unsent data packets are in the send queue when the window closes. The value is 0 when the queue is empty.	UNIX timestamp, in seconds
max_send_success_time	The maximum time that data is successfully sent in the window.	UNIX timestamp, in seconds
send_queue_size	The number of unsent data packets in the current send queue when the window closes.	Number of packets
send_network_error_count	The number of data packets that failed to be sent in the window because of network errors.	Number of packets
send_network_quota_count	The number of data packets that failed to be sent in the window because the quota is exhausted.	Number of packets
send_network_discard_count	The number of discarded data packets in the window because of data exceptions or lack of permissions.	Number of packets
send_success_count	The number of successfully sent data packets in the window.	Number of packets

Keyword	Description	Unit
send_block_flag	Indicates whether the send queue is blocked when the window closes.	N/A
sender_valid_flag	Indicates whether the send flag of this Logstore is valid when the window closes. true indicates that the flag is valid, and false indicates that the flag is invalid due to network or quota errors.	N/A

Logstore statuses

Status	Description	Solution
ok	The status is normal.	No action is required.
process_block	Log parsing is blocked.	Check whether logs are generated at an excessively high speed. If this issue persists, *Configure startup parameters* as needed to modify the upper limit of CPU usage or limits on concurrent inbound traffic to Log Service.
parse_fail	Log parsing failed.	Check whether the log format is consistent with the log collection configuration.
send_block	Current sending is blocked.	Check whether logs are generated at an excessively high speed and whether the network status is normal. If this issue persists, *Configure startup parameters* as needed to modify the upper limit of CPU usage or limits on concurrent inbound traffic to Log Service.

logfile command

Command format

```
/etc/init.d/ilogtaild status logfile [--format={line|json}] index
project-name logstore-name fileFullPath
```

Note:

- **The logfile command is used to query the collection status of a specified log file in LINE or JSON format.**
- **If the `--format=` parameter is not configured, `--format=line` is selected by default. The output is displayed in LINE format.**
- **If this Logstore does not exist or has no log collection activity in the current query window, you get an empty output in LINE format or a `null` value in JSON format.**
- **The `--format` parameter must be placed behind `logfile`.**
- **The `filefullpath` parameter must be a full path name.**

Example

```
/etc/init.d/ilogtaild status logfile 1 sls-zc-test release-test-same /
disk2/test/normal/access.log
time_begin_readable : 17-08-29 11:16:11
time_end_readable : 17-08-29 11:26:11
time_begin : 1503976571
time_end : 1503977171
project : sls-zc-test
logstore : release-test-same
status : ok
config : ##1.0##sls-zc-test$same
file_path : /disk2/test/normal/access.log
file_dev : 64800
file_inode : 22544456
file_size_bytes : 17154060
file_offset_bytes : 17154060
read_bytes : 65033430
parse_success_lines : 230615
parse_fail_lines : 0
last_read_time : 1503977170
read_count : 667
avg_delay_bytes : 0
/etc/init.d/ilogtaild status logfile --format=json 1 sls-zc-test
release-test-same /disk2/test/normal/access.log
{
   "avg_delay_bytes" : 0,
   "config" : "##1.0##sls-zc-test$same",
   "file_dev" : 64800,
   "file_inode" : 22544456,
   "file_path" : "/disk2/test/normal/access.log",
   "file_size_bytes" : 17154060,
   "last_read_time" : 1503977170,
```

```
    "logstore" : "release-test-same",
    "parse_fail_lines" : 0,
    "parse_success_lines" : 230615,
    "project" : "sls-zc-test",
    "read_bytes" : 65033430,
    "read_count" : 667,
    "read_offset_bytes" : 17154060,
    "status" : "ok",
    "time_begin" : 1503976571,
    "time_begin_readable" : "17-08-29 11:16:11",
    "time_end" : 1503977171,
    "time_end_readable" : "17-08-29 11:26:11"
}
```

Output description

Keyword	Description	Unit
status	The collection status of this log file in the current query window. For more information about Logstore status, see the preceding Logstore status table.	N/A
time_begin_readable	The start time of reading.	N/A
time_end_readable	The end time of reading.	N/A
time_begin	The start time of statistics collection.	UNIX timestamp, in seconds
time_end	The end time of statistics collection.	UNIX timestamp, in seconds
project	The project name.	N/A
logstore	The Logstore name.	N/A
file_path	The path of the log file.	N/A
file_dev	The device ID of the log file.	N/A
file_inode	The inode of the log file.	N/A
file_size_bytes	The size of the last scanned file in the window.	Bytes
read_offset_bytes	The parsing offset of this file.	Bytes

Keyword	Description	Unit
config	The collection configuration name. It is a globally unique name which is made up of ##1.0##, project, $, and config.	N/A
read_bytes	The number of logs read in the window.	Bytes
parse_success_lines	The number of successfully parsed log lines in the window.	Line
parse_fail_lines	The number of log lines that failed to be parsed in the window.	Line
last_read_time	The last reading time in the window.	UNIX timestamp, in seconds
read_count	The number of times logs are read in the window.	Number of times
avg_delay_bytes	The average of the differences between the current offset and the file size each time logs are read in the window.	Bytes

history command

Command format

```
/etc/init.d/ilogtaild status history beginIndex endIndex project-name
logstore-name [fileFullPath]
```

Note:

- The history command is used to query the collection status of a Logstore or log file over a period of time.
- `beginIndex` and `endIndex` represent the start and end values for the code query window index, respectively. `beginIndex <= endIndex` is required.

- **If the value of `fileFullPath` is not entered, the collection information of the Logstore is queried. If this value is entered, the collection information of the log files is queried.**

Example

```
/etc/init.d/ilogtaild status history 1 3 sls-zc-test release-test-same
 /disk2/test/normal/access.log
        begin_time          status      read  parse_success
parse_fail      last_read_time  read_count  avg_delay    device
inode  file_size  read_offset
 17-08-29 11:26:11              ok   62.12MB         231000
0   17-08-29 11:36:11         671         0B     64800  22544459    18
.22MB      18.22MB
 17-08-29 11:16:11              ok   62.02MB         230615
0   17-08-29 11:26:10         667         0B     64800  22544456    16
.36MB      16.36MB
 17-08-29 11:06:11              ok   62.12MB         231000
0   17-08-29 11:16:11         687         0B     64800  22544452    14
.46MB      14.46MB
$/etc/init.d/ilogtaild status history 2 5 sls-zc-test release-test-
same
        begin_time          status      read  parse_success
parse_fail      last_read_time  read_count  avg_delay  send_queue
network_error  quota_error  discard_error  send_success  send_block
send_valid          max_unsend          min_unsend    max_send_success
 17-08-29 11:16:11              ok   62.02MB         230615
0   17-08-29 11:26:10         667         0B           0
0            0              0           300       false        true
70-01-01 08:00:00   70-01-01 08:00:00   17-08-29 11:26:08
 17-08-29 11:06:11              ok   62.12MB         231000
0   17-08-29 11:16:11         687         0B           0
0            0              0           303       false        true
70-01-01 08:00:00   70-01-01 08:00:00   17-08-29 11:16:10
 17-08-29 10:56:11              ok   62.02MB         230615
0   17-08-29 11:06:10         687         0B           0
0            0              0           302       false        true
70-01-01 08:00:00   70-01-01 08:00:00   17-08-29 11:06:08
 17-08-29 10:46:11              ok   62.12MB         231000
0   17-08-29 10:56:11         692         0B           0
0            0              0           302       false        true
70-01-01 08:00:00   70-01-01 08:00:00   17-08-29 10:56:10
```

Output description

- **This command is used to query historical collection information of a Logstore or log file in the form of list, with one line for each window.**
- **For more information about each output field, see the description about `logstore` and `logfile` commands.**

Returned values

Normal returned value

The value 0 is returned when all command inputs are valid. The situation where the query on Logstore or log files fails is included. The following content is used as an example:

```
/etc/init.d/ilogtaild status logfile --format=json 1 error-project
error-logstore /no/this/file
null
echo $?
0
/etc/init.d/ilogtaild status all
ok
echo $?
0
```

Abnormal returned value

A non-zero returned value indicates an exception. The following table describes abnormal returned values.

Returned value	Type	Output	Solution
10	Invalid command or lack of parameters	`invalid param, use -h for help.`	Enter `-h` to view the help information.
1	The query time goes beyond the time window ranging from 1 to 60	`invalid query interval`	Enter `-h` to view the help information.
1	Failed to query the specified time window	`query fail, error : $(error)`. For more information, see *errno*.	This issue may occur when the startup time of Logtail is less than the query time span. Submit a ticket if you have any questions.
1	Inconsistency of query time and window time	`no match time interval, please check logtail status`	Check whether Logtail is running . Submit a ticket if you have any questions.

Returned value	Type	Output	Solution
1	No data in the query window	`invalid profile , maybe logtail restart`	Check whether Logtail is running . Submit a ticket if you have any questions.

Example

```
/etc/init.d/ilogtaild status nothiscmd
invalid param, use -h for help.
echo $?
10
/etc/init.d/ilogtaild status/all 99
invalid query interval
echo $?
1
```

Usage scenarios

You can use Logtail health check to understand the overall status of Logtail, and perform collection progress query to obtain related metrics during collection. With the obtained information, you can monitor log collection in a customized manner.

Monitor the running status of Logtail

You can monitor the running status of Logtail by using the `all` command.

How it works: The current status of Logtail is queried every minute. If Logtail is in the `process_block`, `send_block`, or `send_error` state for 5 minutes, an alert is triggered.

You can adjust the alert duration and the range of status to be monitored based on the importance of log collection in specific scenarios.

Monitor the log collection progress

You can monitor the collection progress of a Logstore by using the `logstore` command.

How it works: The `logstore` command is executed every 10 minutes to obtain status information about Logstore. If `avg_delay_bytes` exceeds 1 MB (1024 × 1024) or `status` is not `ok`, an alert is triggered.

The `avg_delay_bytes` alert threshold can be adjusted based on the log collection traffic.

Determine whether collection of a log file is complete

You can determine whether collection of a log file is complete by using the `logfile` command.

How it works: After writing to the log file stops, the `logfile` command is executed every 10 minutes to obtain the status information of this file. If the values for `read_offset_bytes` and `file_size_bytes` are the same, the collection of this log file is complete.

Troubleshoot log collection issues

If log collection is delayed on a server, use the `history` command to query related collection information on this server.

1. **If the value of `send_block_flag` is true, the log collection is delayed due to network issues.**
   - **If the value of `send_network_quota_count` is greater than 0, *split the shard* of the Logstore.**
   - **If the value of `send_network_error_count` is greater than 0, check the network connectivity.**
   - **If no related network error occurs, adjust the *limits on concurrent inbound traffic and throttling* of Logtail.**
2. **The value of `avg_delay_bytes` is greater than a normal one.**
   - **Calculate the average log parsing speed by using `read_bytes` to determine whether the log generation traffic is normal.**
   - **Adjust the *configuration parameters* of Logtail as needed.**
3. **The value of `parse_fail_lines` is greater than 0.**

   Check whether the parsing configurations for log collection takes effect on all logs.

18.7.5.2 Query error information

Errors may occur during Logtail log collection, such as failure to parse regular expressions, incorrect file paths, and traffic exceeding the shard service capabilities. Log Service provides the query function to allow you to query Logtail log collection errors.

Procedure

1. **Go to the Log Collection Error page.**

   Log on to the Log Service console. **Click the specified project to go to the Logstores page. Then, click Diagnose in the Log Collection Mode column. The Log Collection Error page appears.**

2. **View log collection errors.**

   On the Log Collection Error page that appears, you can view the list of Logtail log collection errors of a specified Logstore.

3. **Query log collection errors of a specified machine.**

   To query all log collection errors that occurred on a specific machine, enter the IP address of the machine in the search box on the Log Collection Error page. Logtail reports errors every 5 minutes.

   After an error is rectified, you can check the error time statistics to determine whether the error is reported again after the service recovers. Historical errors are displayed before expiration. You can ignore these errors and check whether any new error is reported after the error is rectified.

Error type	Description	Solution
LOGFILE_PERMINSSION_ALARM	Logtail has no permissions to read the specified file.	Check the Logtail startup account on the server. We recommend that you start Logtail as the root user.
SPLIT_LOG_FAIL_ALARM	The regular expression fails to match with the first line of a log entry and cannot parse the log entry into multiple lines.	Check the correctness of the regular expression to match the first line. For a single line log, you can set the regular expression to `.*`.
MULTI_CONFIG_MATCH_ALARM	A file can be collected by only one Logtail configuration.	Check whether a file is collected by multiple Logtail configurations. If yes, delete the redundant configurations.

Error type	Description	Solution
REGEX_MATCH_ALARM	The log content does not match the regular expression in regular expression mode.	Copy the log sample from the error content for re-matching and generate a new regular expression for parsing.
PARSE_LOG_FAIL_ALARM	Logtail fails to parse logs because the log format does not conform to the definition in the parsing modes such as JSON and delimiter.	Click the error to view the relevant details.
CATEGORY_CONFIG_ALARM	The Logtail collection configuration is invalid.	A common reason is that a regular expression fails to extract the file path as a topic. If this error is caused by other reasons, submit a ticket.
LOGTAIL_CRASH_ALARM	Logtail crashes because it has exceeded the upper limit of machine resource usage.	Modify the upper limits of CPU and memory resources by referring to *Configure startup parameters*. Submit a ticket if you are still unable to resolve this problem.
REGISTER_INOTIFY_FAIL_ALARM	Logtail fails to register with the log listener in Linux possibly because Logtail does not have permissions to access the folder or the folder has been deleted.	Check whether Logtail has the folder access permission and whether the folder exists.

Error type	Description	Solution
DISCARD_DATA_ALARM	This error is caused by insufficient CPU resources configured for Logtail or the network bandwidth throttling.	Modify the upper limit of CPU usage or limits on concurrent inbound traffic to Log Service by following the instructions provided in *Configure startup parameters*. Submit a ticket if you are still unable to resolve this problem.
SEND_DATA_FAIL_ALARM	1. The Apsara Stack tenant account has not created any AccessKey pairs. 2. The Logtail client cannot connect to Log Service, or the network link quality is poor. 3. The write quota of Log Service is insufficient.	1. Check the AccessKey pair. 2. Check the local configuration file /usr/local/ilogtail/ilogtail_config.json, run the curl <service address> command, and check the return result. 3. Add the number of shards for Logstores to support the writing of larger volumes of data.
PARSE_TIME_FAIL_ALARM	Logtail fails to parse the time field based on the time parsing expression.	Configure the time parsing expression correctly based on the log time format.
REGISTER_INOTIFY_FAIL_ALARM	Logtail fails to register with inotify watcher for the log directory.	Check whether the log directory exists. If the directory exists, check the directory permission setting.
SEND_QUOTA_EXCEED_ALARM	The traffic of writing logs exceeds the limit.	*Split shards* in the console.

Error type	Description	Solution
READ_LOG_DELAY_ALARM	Log collection lags behind log generation. Generally, this error is caused by the insufficient CPU resources configured for Logtail or network bandwidth throttling.	Modify the upper limit of the CPU usage or limits on concurrent inbound traffic to Log Service by following the instructions provided in *Query error information*. Submit a ticket if you are still unable to resolve this problem.
DROP_LOG_ALARM	Log collection lags behind log generation, and there are more than 20 unprocessed log rotations. Generally, this error is caused by the insufficient CPU resources configured for Logtail or the network bandwidth throttling.	Modify the upper limit of the CPU usage or limits on concurrent inbound traffic to Log Service by following the instructions provided in *Configure startup parameters*. Submit a ticket if you are still unable to resolve this problem.
LOGDIR_PERMINSSION_ALARM	Logtail has no permission to read the log monitoring directory.	Check whether the log monitoring directory exists. If the directory exists, check the directory permission setting.
ENCODING_CONVERT_ALARM	Log transcoding fails.	Check whether the configured log encoding method is consistent with the actual log encoding.

Error type	Description	Solution
OUTDATED_LOG_ALAR	Logs expire with a time lag of more than 12 hours. Possible causes: Log parsing lags behind by more than 12 hours, the user-defined time field is incorrectly configured, or the time output of the logging program is abnormal.	Check whether READ_LOG_DELAY_ALARM exists. If yes, handle the error according to the instructions for READ_LOG_DELAY_ALARM. If not, check the time field configuration. If the time field is correctly configured, check whether the time output of the logging program is normal. Submit a ticket if you are still unable to resolve this problem.
STAT_LIMIT_ALARM	The number of files in the log collection configuration directory exceeds the upper limit.	Check whether the log collection configuration directory contains a large amount of files and subdirectories, and properly configure the monitored root directory and the maximum monitoring depth of the directory.
DROP_DATA_ALARM	Flushing logs into the local disk times out when the process exits and the unflushed logs are discarded.	Generally, this error is caused by severe collection obstruction. You can modify the upper limit of CPU usage or limits on concurrent inbound traffic to Log Service by following the instructions provided in *Configure startup parameters*. Submit a ticket if you are still unable to resolve this problem.

Error type	Description	Solution
INPUT_COLLECT_ALARM	An exception occurred during the collection at input sources.	Fix the error based on the error message.
HTTP_LOAD_ ADDRESS_ALARM	The entered HTTP address is invalid.	Enter a valid HTTP address.
HTTP_COLLECT_ALARM	An exception occurred during HTTP collection.	Fix the error based on the error message. Typically , this error is caused by timeout.
FILTER_INIT_ALARM	An exception occurred during filter initialization .	Typically, this error is caused by the invalid regular expression of the filter. Fix the error as instructed.
INPUT_CANAL_ALARM	An exception occurred during the running of MySQL binlog.	Fix the error based on the error message. The canal service may be restarted when the configuration is updated. Therefore, you can ignore the service restart error.
CANAL_INVALID_ALARM	The internal status of MySQL binlog is abnormal.	Typically, this error occurs if changes to the table schema cause meta information inconsiste ncy during running. Check whether the table schema is modified when the error is reported. If this error is caused by other reasons, submit a ticket.
MYSQL_INIT_ALARM	An exception occurred during MySQL initializa tion.	Fix the error based on the error message.

Error type	Description	Solution
MYSQL_CHECKPOING_ALARM	The MySQL checkpoint format is invalid.	Determine whether to modify the checkpoint configuration. Submit a ticket if you are still unable to resolve this problem.
MYSQL_TIMEOUT_ALARM	The MySQL query times out.	Check whether the MySQL server and network connection are normal.
MYSQL_PARSE_ALARM	Logtail fails to parse the MySQL query results.	Check whether the checkpoint format configured on MySQL is consistent with the format of the corresponding fields.

 **Note:**

To check all the complete log entries discarded due to a parsing failure, log on to the server and check the `/usr/local/ilogtail/ilogtail.LOG` file.

18.7.5.3 Troubleshoot log collection errors

If an error occurs when you use Logtail to collect logs, perform the following steps for troubleshooting.

Procedure

1. Check whether the account is configured with an AccessKey pair.
   a) *Log on to the Log Service console*.
   b) On the homepage of the console, click the profile picture in the upper right corner. Then, click Personal Information.
   c) On the Personal Information page, click AccessKey. In the dialog box that appears, click OK. Check whether the current account contains an AccessKey pair.

2. **Check whether the Logtail heartbeat of the machine group is normal.**

   Log on to the Log Service console. **On the Machine Groups page, check the status of the machine group. If the heartbeat status is OK, perform the next step; if the heartbeat status is FAIL, proceed with the troubleshooting.**

   Generally, Logtail heartbeat failures are caused by the following reasons:

   - **Logtail is not installed**

     Check the client status in Linux:

     ```
     sudo /etc/init.d/ilogtaild status
     ```

     If the Logtail client is not installed, see *Install Logtail in Linux* **to install Logtail on the server for log collection.**

   - **Incorrect parameters are configured during installation**

     As Log Service operates by region, you must specify the correct server-end endpoint when installing the Logtail client. Check the configuration of the installed client in the following paths:

     For Linux: `/usr/local/ilogtail/ilogtail_config.json`

     Perform the following operations:

     - **Verify that the endpoint specified for the client is in the same region as the Log Service project.**
     - **Verify that the correct domain is selected based on the network environment of machines. If an internal domain is selected for a machine in a VPC, the client connection will fail. You can use Telnet to test the domain**

configured in *ilogtail_config.json*. The following content is used as an example: `telnet logtail.cn-hangzhou-intranet.log.aliyuncs.com 80`.

- **An incorrect IP address or user-defined identifier is configured on the server**

  Generally, a Logtail client obtains the IP address of a machine in one of the following ways:

  - **If hostname binding is configured in the *`/etc/hosts`* file, confirm the bound IP address. Run the `hostname` command to view the host name.**
  - **If no hostname is bound, the Logtail client obtains the IP address of the first NIC on the local machine.**

  Check the IP address on the server in the following paths:

  For Linux: *`/usr/local/ilogtail/app_info.json`*

  If the IP address entered in the machine group at the server end is different from that obtained by the Logtail client, make adjustments based on the actual situation:

  - **If an incorrect IP address is entered in the machine group at the server end , modify the IP address and save it. Wait 1 minute and check the IP address consistency again.**
  - **If the network configuration (for example, *`/etc/hosts`*) of the machine is modified, restart the Logtail client to obtain a new IP address.**

  Run the following command or perform the following operations to restart the Logtail client as needed:

  For Linux: `sudo /etc/init.d/ilogtaild stop; sudo /etc/init.d/ilogtaild start`

3. **Check whether the collection configuration is created and applied to the machine group.**

   After confirming that the status of the Logtail client is normal, check the following configurations:

   a) **Check whether the Logtail configuration is created.**

      Verify that the monitored directory and log file name are the same as those on the machine. The directory structure supports both the complete path mode and the wildcard mode.

   b) **Verify that the Logtail configuration has been applied to the machine group.**

      Check the Logtail machine group. Select Config, and check whether the target configuration is applied to the machine group.

4. **Check collection errors.**

   If Logtail is properly configured, check whether new data is generated in real time in the log file. As Logtail collects only incremental data, it does not read inventory files that are not updated. If a log file is updated but the updates cannot be queried in Log Service, you can identify the reasons of the problem using the following methods:

   - **View Logtail logs**

     Client logs include key information and all warning and error logs. To query complete and real-time error information, view client logs in the following paths:

     For Linux: `/usr/local/ilogtail/ilogtail.LOG`

   - **Check whether the usage exceeds the limit**

     To collect large volumes of logs or files, you can modify the Logtail startup parameters to obtain higher log collection throughput. For more information about how to make adjustments, see *Configure startup parameters*.

   If the problem persists, submit a ticket to Log Service engineers and attach key information collected during troubleshooting.

18.7.6 Limits

This topic describes the limits on Logtail, including limits on file collection, resources, and error handling.

Table 18-3: File collection limits

Type	Limit
File encoding	Logs encoded in UTF-8 and GBK are supported. We recommend that you use UTF-8 to improve processing performance. Log files encoded in other formats result in unexpected behaviors such as gibberish and data loss.
Log file size	Unlimited.
Log file rotation	The `.log*` and `.log` files are supported.
Log collection behavior upon log parsing block	When log parsing is blocked, Logtail retains the open state of the log file descriptor (FD). If log file rotation occurs multiple times during the block, Logtail attempts to keep the parsing sequence of each rotated log file. If more than 20 rotated log files need to be parsed, Logtail does not process subsequent log files.
Soft link	Monitored directories can be soft links.
Size of a single log	The maximum size of a single log is 512 KB. If a multi-line log is divided by using a regular expression to match the first line, the maximum size of each log after division is still 512 KB. If the size of a log exceeds 512 KB, the log is forcibly split into multiple parts for collection. For example, if a log is 1025 KB, it will be split into three parts: 512 KB, 512 KB, and 1 KB. These log parts are collected consecutively.
Regular expression	Regular expressions can be Perl-compatible regular expressions.

Type	Limit
Applying multiple collection configurations to the same file	Not supported. We recommend that you collect log files to one Logstore and configure multiple subscriptions. If this feature is required, configure soft links for log files to bypass the limit.
File opening behavior	Logtail retains the open state of a file to be collected. Logtail closes the file if the file is not modified for more than 5 minutes (in case that rotation does not occur).
First log collection behavior	Logtail collects only incremental log files. When a log file is modified for the first time and the size of the log file exceeds 1 MB, Logtail collects logs from the last 1 MB of the log file. If the size of the log file is no greater than 1 MB, Logtail collects logs from the beginning of the log file. If a log file is not modified after the configuration is pushed, Logtail does not collect this log file.
Non-standard text log	For a log that contains the characters \0, the log is truncated to the position where the characters \0 first appear.

Table 18-4: Checkpoint management

Item	Description
Checkpoint timeout interval	If a file remains unmodified for more than 30 days, the checkpoint is deleted.
Checkpoint saving policy	Checkpoints are saved every 15 minutes , and are automatically saved when the program exits.
Checkpoint saving path	The default saving path is `/tmp/logtail_checkpoint`. You can modify the parameters according to *Configure startup parameters*.

Table 18-5: Configuration limits

Item	Description
Configuration update	A custom configuration update takes effect after about 30 seconds.
Dynamic configuration loading	Supported. Updates on a collection configuration do not affect other collection configurations.
Number of collection configuration files	Unlimited. However, we recommend that the number of collection configuration files for a server be no more than 100.
Multi-tenant isolation	Collection configurations for different tenants are isolated.

Table 18-6: Resources and performance limits

Item	Description
Log processing throughput	The default throughput in processing raw logs is limited to 2 MB/s. (Data is uploaded after encoding and compression, with a general compression ratio of 5:1 to 10:1.) Logs may be lost if the processing throughput is exceeded. You can modify the parameters according to *Configure startup parameters*.
Maximum performance	Single-core capability: The maximum processing capability is 100 MB/s for logs in simple format, 20 MB/s for logs in regular expression format (the capability varies with the complexity of regular expressions), 40 MB/s for logs in delimiter-separated format, and 30 MB/s for logs in JSON format. After multiple processing threads are enabled, the performance can be improved by 1.5 to 3 times.
Number of monitored directories	Logtail proactively limits the number of monitored directories to reduce the consumption of resources. If the upper limit is reached, Logtail stops monitoring any more directories and log files. Logtail monitors a maximum of 3,000 directories, including subdirectories.
Default limits on resources	By default, Logtail occupies up to 40% of CPU and 256 MB of memory. If logs are generated at a high speed, you can modify the parameters according to *Configure startup parameters*.

Item	Description
Processing policy for resource limit exceeding	If the resources occupied by Logtail exceed the upper limit and this issue lasts for 3 minutes, Logtail is forced to restart. The restart may cause loss or duplication of data.

Table 18-7: Error handling limits

Item	Description
Network error handling	If a network error occurs, Logtail proactively retries and automatically adjusts the retry interval.
Handling of resource limit exceeding	If the data transmission rate exceeds the upper limit of Logstore, Logtail blocks log collection and automatically retries.
Maximum retry period in the case of timeout	If data fails to be transmitted for more than 6 successive hours, Logtail discards the data.
Status self-check	Logtail automatically restarts in the case of an exception, for example, abnormal exit of a program or resource limit exceeding.

Table 18-8: Other limits

Item	Description
Log collection delay	Normally, the delay in log collection by Logtail does not exceed one second after logs are flushed to a disk. This does not apply when log collection is blocked.
Log upload policy	Logtail automatically aggregates logs in the same file before uploading the logs. Log uploading is triggered if the number of logs exceeds 2,000, the total size of logs exceeds 2 MB, or the log collection duration exceeds 3 seconds.

18.8 Other collection methods

18.8.1 Logstash

18.8.1.1 Logstash overview

Log Service supports collection of server logs through Logstash and upload of data to Log Service through a plug-in.

Log Service supports collection of logs through APIs, SDKs, and Logstash. As an open source log management tool, Logstash can collect and process distributed and diversified logs, and transfer them to a specified location, for example, a server or a file. You can install Logstash and plug-ins on ECS instances, on-premises machines, or virtual machines of other cloud service vendors. After simple configuration, you can migrate server logs to the cloud.

After you install Logstash and its related plug-ins on machines, and configure file directories, Log Service projects, and Log Service Logstores, Logstash automatically traces the changes of log files, collects log files in real time, parses the log files, and sends them to Log Service.

Log Service supports the input of data through Logstash. Logstash provides the following features:

- **Collect logs of various types on machines and support data sources such as files, TCP, and syslog.**
- **Interconnect with the account security system and support data signature transfer and access control by using AccessKey pairs.**
- **Support batch transfer of logs to reduce the TPS costs in writing data into Log Service.**
- **Compress log data and then transfer the data to Log Service to reduce the occupation of network egress bandwidth.**

18.8.1.2 Quick installation

To quickly install Logstash on your server, you can choose the default installation mode.

Context

Log Service provides an installation package based on Logstash 2.2.2. The package integrates with JRE 1.8, Log Service output plug-in, and NSSM 2.24. Compared with

custom installation, installing Logstash by using this package is more convenient. For advanced requirements, you can choose the custom installation mode.

Procedure

1. Download the *installation package* and decompress it to the C: drive.
2. Verify that the path for the Logstash startup program is `C:\logstash-2.2.2-win\bin\logstash.bat`.

18.8.1.3 Custom installation

You can perform custom installation when installing Logstash.

Context

Custom installation allows you to perform parameter configurations, such as the modification of default settings.

Procedure

1. Install Java.
   a. Download the installation package.

      Visit the *Java official website* to download JDK, and then double click the JDK to install Java.

   b. Set environment variables.

      Add or modify environment variables in advanced system settings.

      - **PATH:** `C:\Program Files\Java\jdk1.8.0_73\bin`
      - **CLASSPATH:** `C:\Program Files\Java\jdk1.8.0_73\lib;C:\Program Files\Java\jdk1.8.0_73\lib\tools.jar`
      - **JAVA_HOME:** `C:\Program Files\Java\jdk1.8.0_73`

   c. Perform verification.

      Use `PowerShell` or `cmd.exe` for verification.

```
PS C:\Users\Administrator> java -version
java version "1.8.0_73"
Java(TM) SE Runtime Environment (build 1.8.0_73-b02)
Java HotSpot(TM) 64-Bit Server VM (build 25.73-b02, mixed mode)
PS C:\Users\Administrator> javac -version
```

```
javac 1.8.0_73
```

2. **Install Logstash.**

   a. **Download the installation package from the official website.**

      Select version 2.2 or later on the *Logstash* home page.

   b. **Install Logstash.**

      Decompress *logstash-2.2.2.zip* to the *C:\logstash-2.2.2* directory.

      Verify that the path of Logstash startup program is *C:\logstash-2.2.2\bin\logstash.bat*.

3. **Install the plug-in used by Logstash to write logs to Log Service.**

   Choose the plug-in installation method based on the network environment where the machine resides.

   - **Online installation**

     The plug-in is hosted by RubyGems. For more information, see *the related description*.

     Use `PowerShell` or `cmd.exe` to go to the Logstash installation directory. Run the following commands to install Logstash:

     ```
     PS C:\logstash-2.2.2> .\bin\plugin install logstash-output-logservice
     ```

   - **Offline installation**

     Download from the official website: Access the *logstash-output-logservice page* and click Download in the lower-right corner.

     If the machine from which logs are collected cannot access the Internet, copy the downloaded gem package to the *C:\logstash-2.2.2* directory of

the machine. Use *PowerShell* **or** *cmd.exe* **to go to the Logstash installation directory. Run the following commands to install Logstash:**

```
PS C:\logstash-2.2.2> .\bin\plugin install C:\logstash-2.2.2\
logstash-output-logservice-0.2.0.gem
```

- **Verification**

```
PS C:\logstash-2.2.2> .\bin\plugin list
```

Verify that logstash-output-logservice exists in the list of installed plug-ins on the machine.

4. **Install NSSM.**

   Download from the official website: Visit the *NSSM official website* **to download the NSSM installation package.**

   Decompress the downloaded NSSM installation package to the *C:\logstash-2.2.2\nssm-2.24* **directory.**

18.8.1.4 Set Logstash to a Windows service

To facilitate automatic log collection, you can set Logstash to a Windows service so that Logstash can work in the back end and start automatically after power-on.

Context

If you start *logstash.bat* **in PowerShell, the Logstash process will run in the front end. This configuration is usually used for configuration testing and log collection debugging. We recommend that you set Logstash to a Windows service after debugging so that Logstash can work in the back end and start automatically after power-on.**

Besides setting Logstash to a Windows service, you can use command lines to start, stop, modify, and delete services. For more information about how to use NSSM, see *official NSSM document***.**

Add a service

This operation is generally performed when Logstash is deployed for the first time. If Logstash has been added, skip this step.

You can run the following commands to add Logstash as a Windows service.

- **32-bit system**

```
C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe install logstash "C:\
logstash-2.2.2-win\bin\logstash.bat" "agent -f C:\logstash-2.2.2-win
\conf"
```

- **64-bit system**

```
C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe install logstash "C:\
logstash-2.2.2-win\bin\logstash.bat" "agent -f C:\logstash-2.2.2-win
\conf"
```

Start the service

If a configuration file in the Logstash *conf* directory is updated, stop the Logstash service and then start it again.

You can run the following commands to start the service.

- **32-bit system**

```
C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe start logstash
```

- **64-bit system**

```
C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe start logstash
```

Stop the service

You can run the following commands to stop the service.

- **32-bit system**

```
C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe stop logstash
```

- **64-bit system**

```
C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe stop logstash
```

Modify the service

You can run the following commands to modify the service.

- **32-bit system**

```
C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe edit logstash
```

- **64-bit system**

```
C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe edit logstash
```

Delete the service

You can run the following command to delete the service.

- **32-bit system**

```
C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe remove logstash
```

- **64-bit system**

```
C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe remove logstash
```

18.8.1.5 Create a Logstash collection configuration

Log Service supports the collection of server logs through Logstash. After you configure Logstash collection for a data source, Logstash pushes logs to Log Service in real time.

Related plug-ins

- **logstash-input-file**

  This plug-in is used to collect log files in tail mode. For more information, see *logstash-input-file*.

  Note:
  path indicates the file path, which must use UNIX delimiters. Otherwise, fuzzy matching is not supported. A sample file path is as follows: `C:/test/multiline/*.log`.

- **logstash-output-logservice**

  This plug-in is used to send the logs collected by the logstash-input-file plug-in to Log Service.

Parameter	Description
endpoint	**The endpoint of Log Service. Example: `http://regionid.example.com`.**
project	**The name of a Log Service project.**

Parameter	Description
logstore	The name of a Logstore.
topic	The topic of logs. The default value is null.
source	The log source. If the value of this parameter is null, the IP address of the current machine is used as the log source. Otherwise, the log source is the specified parameter value.
access_key_id	The AccessKey ID of an account.
access_key_secret	The AccessKey secret of an account.
max_send_retry	The maximum number of retries performed when data packets fail to be sent to Log Service. Data packets that fail to be sent within the retry period are discarded. The retry interval is 200 ms.

Procedure

1. **Create a collection configuration.**

   Create a configuration file in the `C:\logstash-2.2.2-win\conf\` directory and then restart Logstash to make the file take effect.

   You can create a configuration file for each type of logs. The file name format is `*.conf`. To facilitate management, we recommend that you create all the configuration files in the `C:\logstash-2.2.2-win\conf\` directory.

    **Note:**

Configuration files must be encoded in UTF-8 without BOM format. You can use Notepad++ to modify the file encoding format.

- **IIS logs**

  See *Use Logstash to collect IIS logs*.

- **CSV logs**

  The system time when the logs are collected is used as the log upload time. For more information, see *Use Logstash to collect CSV logs.*

- **Logs with built-in time**

  Take the format of CSV logs as an example. The time in the log content is used as the log upload time. For more information, see *Use Logstash to collect CSV logs*.

- **General logs**

  By default, the system time when the logs are collected is used as the log upload time. Log fields are not parsed. Single-line logs and multi-line logs are supported. For more information, see *Use Logstash to collect other logs*.

2. **Verify the configuration syntax.**

   a. **Use** `PowerShell` **or** `cmd.exe` **to go to the Logstash installation directory. Run the following commands to verify the configuration:**

   ```
   PS C:\logstash-2.2.2-win\bin> .\logstash.bat agent --configtest --
   config C:\logstash-2.2.2-win\conf\iis_log.conf
   ```

   b. **Modify the collection configuration file. Add the temporary configuration item rubydebug in the output phase to send the collected results to the console. Set the type field as needed.**

   ```
   output {
   if [type] == "***" {
     stdout { codec => rubydebug }
     logservice {
     ...
     }
    }
   ```

```
}
```

c. Use `PowerShell` or `cmd.exe` to go to the Logstash installation directory to start the process. Run the following commands:

```
PS C:\logstash-2.2.2-win\bin> .\logstash.bat agent -f C:\logstash-2.2.2-win\conf
```

After the verification, end the *logstash.bat* process and delete the temporary configuration item rubydebug.

What's next

If you start *logstash.bat* in PowerShell, the Logstash process will run in the front end. This configuration is usually used for configuration testing and log collection debugging. We recommend that you set Logstash to a Windows service after debugging so that Logstash can work in the back end and start automatically after power-on. For more information about how to set Logstash to a Windows service, see *Set Logstash to a Windows service*.

18.8.1.6 Advanced functions

Log Service supports the collection of server logs through Logstash and upload of data to Log Service through plug-ins.

Logstash provides *multiple plug-ins* to meet personalized requirements. Some plug-ins are described as follows:

- *grok*: Structurally parses logs into multiple fields by using regular expressions.
- *json_lines* and *json*: Structurally parse JSON logs.
- *date*: Parses and converts the date and time fields of logs.
- *multiline*: Customizes complex types of multi-line logs.
- *kv*: Structurally parses logs of key-value pair type.

18.8.1.7 Logstash error handling

If you encounter the following collection errors when using Logstash to collect logs, follow the following suggestions to troubleshoot the errors.

- Garbled characters displayed in Log Service

  Logstash supports UTF-8 file encoding by default. Check whether input files are correctly encoded.

- **Error message displayed in the console**

  When the error message `io/console not supported; tty will not be manipulated` **is prompted in the console, you do not need to handle it because the error does not affect functions.**

  If other errors occur, we recommend that you search Google or Logstash forums for help.

18.8.2 SDK collection

18.8.2.1 Producer Library

LogHub Producer Library is a LogHub class library for writing highly concurrent Java application data. Producer Library and Consumer Library are used for LogHub read/write encapsulation, lowering the threshold for data collection and consumption.

Features

- **Provides an asynchronous sending interface, ensuring thread security.**
- **Allows you to add configurations of multiple projects.**
- **Supports the configuration of the quantity of network I/O threads used for data transmission.**
- **Supports the configuration of the quantity and size of logs merged into a package .**
- **Supports controllable memory usage. When the memory usage reaches the configured threshold, the Send interface of Producer will be blocked until there is idle memory.**

Benefits

- **Logs collected from clients are not flushed into the disk. The collected logs are directly sent to the LogHub server through the network.**
- **Highly concurrent write operations on clients are achieved. For example, more than one hundred write operations are generated within 1 second.**
- **Computing and I/O on clients are logically separated. Logging does not affect the time spent on computing.**

In the preceding scenarios, Producer Library can simplify program development , aggregate write requests in batches, and send the requests to the LogHub server

asynchronously. During the process, you can configure the parameters for batch aggregation and the logic to process server exceptions.

Comparison of different access modes:

Access mode	Advantage/Disadvantage	Target scenario
SDK direct transmission	Logs are not flushed into a disk. They are directly sent to the server. Switching between the network I/O and program I/O needs to be properly processed.	Scenarios where logs are not flushed into a disk
Producer Library	Logs are not flushed into a disk. Write requests are asynchronously aggregated and sent to the server. High throughput is required.	Scenarios where logs are not flushed into a disk and the QPS on clients is high

Procedure

- *Java Producer*
- *Log4J1.XAppender (based on Java Producer)*
- *Log4J2.XAppender (based on Java Producer)*
- *LogBack Appender (based on Java Producer)*

18.8.2.2 Log4j Appender

Alibaba Cloud Log Log4j Appender enables you to set the log output destination to Log Service.

Loghub Log4j Appender

Log4j is an open-source project of Apache. Log4j allows you to set the log output destination to console, file, GUI component, socket server, NT event recorder, or UNIX Syslog daemon. You can also set the output format of each log, and define the level of each log to implement refined log generation control. These features can be achieved through a configuration file without the need to modify the code of applications.

For more information about the download link and user guide, see *Github*.

18.8.2.3 C Producer Library

Besides the Java Producer Library, LogHub also supports the C Producer Library and C Producer Lite Library. LogHub provides a cross-platform end-to-end log

collection solution that features simplicity, high performance, and low resource consumption.

18.8.3 Common log formats

18.8.3.1 Overview

The following topics list the common log formats and provide the methods for configuring Logtail to collect such logs.

18.8.3.2 Apache logs

Typically, the formats and directories of Apache logs are defined or specified in the configuration file *`/etc/apache2/httpd.conf`*.

Log formats

The Apache log configuration file defines two log formats: combined and common.

- **Combined:**

```
LogFormat "%h %l %u %t \"%r\" %>s %b \"%{Referer}i\" \"%{User-Agent}
i\"" combined
```

- **Common:**

```
LogFormat "%h %l %u %t \"%r\" %>s %b"
```

The following statement uses the combined format and the specified file name.

```
CustomLog "/var/log/apache2/access_log" combined
```

Field description

Field format	Meaning
%a	remote_ip
%A	local_ip
%B	size
%b	size
%D	time_taken_ms
%h	remote_host
%H	protocol
%l	ident
%m	method

Field format	Meaning
%p	port
%P	pid
“%q”	url_query
“%r”	request
%s	status
%>s	status
%t	time
%T	time_taken
%u	remote_user
%U	url_stem
%v	server_name
%V	canonical_name
%I	bytes_received
%O	bytes_sent
“%{User-Agent}i”	user_agent
“%{Referer}i”	referer

Sample log

```
192.168.1.2 - - [02/Feb/2016:17:44:13 +0800] "GET /favicon.ico HTTP/1.
1" 404 209 "http://localhost/x1.html" "Mozilla/5.0 (Macintosh; Intel
Mac OS X 10_11_3) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/48.0.
2564.97 Safari/537.36"
```

Configure Logtail to collect Apache logs

1. **Create a project and a Logstore. For the detailed procedures, see** *Create a project* **and** *Create a Logstore*.
2. **On the Logstores page, click the Data Import Wizard icon.**
3. **Select a data type.**

   Select Text File and click Next.

4. **Configure a data source.**

   a. **Enter the configuration name and log path. Then, set the log collection mode to Full Regex Mode.**

   b. **Enter a sample log and enable Extract Field.**

   c. **Highlight fields to generate a regular expression, and manually adjust it.**

      Log Service can automatically parse the highlighted fields of the sample log to generate a regular expression. There may be minor differences in the actual log formats. You can click Manually Input Regular Expression to adjust the automatically generated regular expression. This makes the regular expression suitable for all formats of the logs collected.

      Click Validate after modifying the regular expression. If the regular expression is correct, extraction results are displayed. If any error occurs, adjust the regular expression.

   d. **Enter the corresponding keys for log content extraction results.**

      Specify a descriptive field name for each log field extraction result. For example, specify time for the time field. Enable Use System Time and then click Next.

   After configuring Logtail, apply the configuration to the machine group to start collecting Apache logs.

18.8.3.3 NGINX log

The NGINX log format and directory are generally specified in the *`/etc/nginx/nginx.conf`* **configuration file.**

NGINX log format

The log configuration file defines the print format of NGINX logs, that is, the main format:

```
log_format main  '$remote_addr - $remote_user [$time_local] "$request"
 '
                 '$request_time $request_length '
                 '$status $body_bytes_sent "$http_referer" '
```

```
                  '"$http_user_agent"';
```

The following code declares that the main log format is used and declares the name of the file to write logs to.

```
access_log /var/logs/nginx/access.log main
```

NGINX log fields

Field	Description
remoteaddr	The IP address of the client.
remote_user	The username of the client.
request	The requested URL and HTTP.
status	The request status.
bodybytessent	The number of bytes sent to the client excluding the size of the response header. The value of this variable is the same as the value of bytes_sent in modlogconfig of the Apache module.
connection	The serial number of a connection.
connection_requests	The number of requests received by using a connection.
msec	The time when the log is written, in seconds, with millisecond precision.
pipe	Indicates whether requests are sent by using the HTTP pipeline. When requests are sent by using the HTTP pipeline, the value is p. Otherwise, the value is a period (.).
httpreferer	The source page of the access request.
“http_user_agent”	The browser information of the client , which must be enclosed by double quotation marks (").
requestlength	The request length, which includes the request line, request header, and request body.

Field	Description
request_time	The period of time when the request is processed, in seconds, with millisecond precision. The time starts when the first character is sent to the client and ends when the logs are written after the last character is sent to the client.
[$time_local]	The local time when the general log format is applied, which must be enclosed by brackets ([ ]).

Sample log

```
192.168.1.2 - - [10/Jul/2015:15:51:09 +0800] "GET /ubuntu.iso HTTP/1.0
" 0.000 129 404 168 "-" "Wget/1.11.4 Red Hat modified"
```

Configure Logtail to collect NGINX logs

1. Create a project and a Logstore. For more information, see *Create a project* and *Create a Logstore*.
2. On the Logstores page, click the icon in the Data Import Wizard column.
3. Select a data type.

   Select NGINX Access Log and click Next.
4. Configure a data source.

   a. Enter the values of Configuration Name and Log Path.

   b. Enter the NGINX log format.

      Enter the section of log configurations in a standard NGINX configuration file. The value typically starts with `log_format`. Log Service automatically reads the NGINX key.

   c. Set Advanced Options as needed, and then click Next.

      For more information about advanced options, see *Advanced options*.

After configuring Logtail, apply the configuration to the machine group to collect NGINX logs.

18.8.3.4 Python log

The Python logging module provides a general logging system, which can be used by third-party modules or applications. The logging module defines different log

levels and records logs by using different methods, such as file, HTTP GET/POST, SMTP, and Socket. Additionally, the logging module allows you to customize the method of recording logs. The logging module uses the same mechanism as Log4j except for different implementation details. The logging module provides the logger, handler, filter, and formatter features.

Python log format

The formatter specifies the output format of log records. Formatter uses two parameters to specify the log format: message format string and message date string. Both of the parameters are optional.

Python log format:

```
import logging
import logging.handlers
LOG_FILE = 'tst.log'
handler = logging.handlers.RotatingFileHandler(LOG_FILE, maxBytes =
1024*1024, backupCount = 5) # Instantiate the handler
fmt = '%(asctime)s - %(filename)s:%(lineno)s - %(name)s - %(message)s'

formatter = logging.Formatter(fmt)   # Instantiate the formatter
handler.setFormatter(formatter)      # Add the formatter to the
handler
logger = logging.getLogger('tst')    # Obtain the logger named tst
logger.addHandler(handler)           # Add the handler to the logger
logger.setLevel(logging.DEBUG)
logger.info('first info message')
logger.debug('first debug message')
```

Field description

The formatter is configured in `%(key)s` format, that is, replacing dictionary keywords. The following table lists the provided keywords.

Format	Description
%(name)s	The name of the logger that generates logs.
%(levelno)s	The log level in numeric format, including the numbers representing the levels of DEBUG, INFO, WARNING, ERROR, and CRITICAL.
%(levelname)s	The log level in text format, including 'DEBUG', 'INFO', 'WARNING', 'ERROR', and 'CRITICAL'.

Format	Description
%(pathname)s	The full path of the source file where the statement that generates the log resides (if available).
%(filename)s	The file name.
%(module)s	The name of the module where the statement that generates the log resides.
%(funcName)s	The name of the function that calls the log output function.
%(lineno)d	The line of code where the statement that calls the log output function resides (if available).
%(created)f	The time when the log is created, in UNIX time format, which indicates the number of seconds that have elapsed since January 1, 1970 00:00:00 (UTC).
%(relativeCreated)d	The interval between the time when the log is created and the time when the logging module is loaded, accurate to milliseconds.
%(asctime)s	The log creation time. The value of 2003-07-08 16:49:45,896 shows the default format. The number after the comma (,) indicates the number of milliseconds.
%(msecs)d	The time when the log is created, accurate to milliseconds.
%(thread)d	The thread ID (if available).
%(threadName)s	The thread name (if available).
%(process)d	The process ID (if available).
%(message)s	The log message.

Sample log

Sample output log:

```
2015-03-04 23:21:59,682 - log_test.py:16 - tst - first info message
```

```
2015-03-04 23:21:59,682 - log_test.py:17 - tst - first debug message
```

Configure Logtail to collect Python logs

For more information about how to configure Logtail to collect Python logs, see *Apache logs*. Select the corresponding configuration based on the network deployment and the actual situation.

The automatically generated regular expression is based on the sample log only and may not be applicable to other logs. Therefore, you need to adjust the regular expression after it is automatically generated.

Common Python logs and the corresponding regular expressions:

- **Sample log:**

  ```
  2016-02-19 11:03:13,410 - test.py:19 - tst - first debug message
  ```

  Regular expression:

  ```
  (\d+-\d+-\d+\s\S+)\s+-\s+([^:]+):(\d+)\s+-\s+(\w+)\s+-\s+(. *)
  ```

- **Log format:**

  ```
  %(asctime)s - %(filename)s:%(lineno)s - %(levelno)s %(levelname)
  s %(pathname)s %(module)s %(funcName)s %(created)f %(thread)d %(
  threadName)s %(process)d %(name)s - %(message)s
  ```

  Sample log:

  ```
  2016-02-19 11:06:52,514 - test.py:19 - 10 DEBUG test.py test <module
  > 1455851212.514271 139865996687072 MainThread 20193 tst - first
  debug message
  ```

  Regular expression:

  ```
  (\d+-\d+-\d+\s\S+)\s-\s([^:]+):(\d+)\s+-\s+(\d+)\s+(\w+)\s+(\S+)\s
  +(\w+)\s+(\S+)\s+(\S+)\s+(\d+)\s+(\w+)\s+(\d+)\s+(\w+)\s+-\s+(. *)
  ```

18.8.3.5 Log4j log

Log4j logs include Log4j 1 logs and Log4j 2 logs. This topic describes how to configure regular expressions based on the default configuration of Log4j 1 logs.

Access methods

Log Service can collect Log4j logs through the following methods:

- **LogHub Log4j Appender**
- **Logtail**

Collect Log4j logs through Loghub Log4j Appender

For more information, see *Log4j Appender*.

Collect Log4j logs through Logtail

This topic describes how to configure regular expressions based on the default configuration of Log4j 1 logs. If Log4j 2 is used, you must modify the default configuration to record the complete date information.

```
<Configuration status="WARN">
  <Appenders>
    <Console name="Console" target="SYSTEM_OUT">
      <PatternLayout pattern="%d{yyyy-MM-dd HH:mm:ss:SSS zzz} [%t] %-
5level %logger{36} - %msg%n"/>
    </Console>
  </Appenders>
  <Loggers>
    <Logger name="com.foo.Bar" level="trace">
      <AppenderRef ref="Console"/>
    </Logger>
    <Root level="error">
      <AppenderRef ref="Console"/>
    </Root>
  </Loggers>
</Configuration>
```

For more information about how to configure Logtail to collect Log4j logs, see *Apache logs*. Select the corresponding configuration based on your network deployment and actual situation.

The automatically generated regular expression is based only on the sample log and may not be applicable to other logs. Therefore, you need to adjust the automatically generated regular expression.

The following sample log is a Log4j log that is written to a file based on the default log format:

```
2013-12-25 19:57:06,954 [10.207.37.161] WARN impl.PermanentTairDaoImpl
 - Fail to Read Permanent Tair,key:e:470217319319741_1,result:com
```

```
.example.tair.Result@172e3ebc[rc=code=-1, msg=connection error or
timeout,value=,flag=0]
```

Matching of the first line for a multi-line log (with an IP address indicating the beginning of a line):

```
\d+-\d+-\d+\s. *
```

Regular expression used to extract log information:

```
(\d+-\d+-\d+\s\d+:\d+:\d+,\d+)\s\[([^\]]*)\]\s(\S+)\s+(\S+)\s-\s(. *)
```

Time conversion format:

```
%Y-%m-%d %H:%M:%S
```

The following table describes extraction results of the sample log.

Key	Value
time	**2013-12-25 19:57:06,954**
ip	**10.207.37.161**
level	**WARN**
class	**impl.PermanentTairDaoImpl**
message	**Fail to Read Permanent Tair,key:e:470217319319741_1,result :com.example.tair.Result@172e3ebc[rc=code=-1, msg= connection error or timeout,value=,flag=0]**

18.8.3.6 Node.js log

Node.js logs are displayed in the console by default, which is inconvenient for data collection and troubleshooting. With Log4js, logs can be written to files and the log format can be customized, facilitating data collection and consolidation.

```
var log4js = require('log4js');
log4js.configure({
  appenders: [
    {
      type: 'file', //File output
      filename: 'logs/access.log',
      maxLogSize: 1024,
      backups:3,
      category: 'normal'
    }
  ]
});
var logger = log4js.getLogger('normal');
logger.setLevel('INFO');
logger.info("this is a info msg");
```

```
logger.error("this is a err msg");
```

Log formats

After logs are stored into text files by using Log4js, logs are displayed in the following format in the files:

```
[2016-02-24 17:42:38.946] [INFO] normal - this is a info msg
[2016-02-24 17:42:38.951] [ERROR] normal - this is a err msg
```

Log4js defines six log output levels, including trace, debug, info, warn, error, and fatal in ascending order.

Use Logtail to collect Node.js logs

For more information about how to configure Logtail to collect Node.js logs, see *Apache logs*. Select the corresponding configuration based on the network deployment and the actual situation.

The automatically generated regular expression is based on the sample log only and may not be applicable to other logs. Therefore, you need to adjust the regular expression after it is automatically generated. You can refer to the following sample Node.js logs to configure appropriate regular expressions for your logs.

Common Node.js logs and the corresponding regular expressions:

- **Sample Node.js log 1:**
  - **Sample log:**

    ```
    [2016-02-24 17:42:38.946] [INFO] normal - this is a info msg
    ```

  - **Regular expression:**

    ```
    \[([^]]+)]\s\[([^\]]+)]\s(\w+)\s-(. *)
    ```

  - **Extracted fields:**

    `time`, `level`, `loggerName`, **and** `message`

- **Sample Node.js log 2:**
  - **Sample log:**

    ```
    [2016-01-31 12:02:25.844] [INFO] access - 42.120.73.203 - - "GET /
    user/projects/ali_sls_log? ignoreError=true HTTP/1.1" 304 - "http
    ://
    ```

```
aliyun.com/" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_3)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/48.0.2564.97 Safari/
537.36"
```

- **Regular expression:**

```
\[([^]]+)]\s\[(\w+)]\s(\w+)\s-\s(\S+)\s-\s-\s"([^"]+)"\s(\d+)
[^"]+("[^"]+)"\s"([^"]+). *
```

- **Extracted fields:**

`time`, `level`, `loggerName`, `ip`, `request`, `status`, `referer`, **and** `user_agent`

18.8.3.7 WordPress log

This topic describes the default format of WordPress logs to provide you with a reference for configuring Logtail to collect WordPress logs.

Default format of WordPress logs

Sample raw log:

```
172.64.0.2 - - [07/Jan/2016:21:06:39 +0800] "GET /wp-admin/js/password
-strength-meter.min.js? ver=4.4 HTTP/1.0" 200 776 "http://wordpress
.c4a1a0aecdb1943169555231dcc4adfb7.cn-hangzhou.alicontainer.com/wp-
admin/install.php" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_5)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/47.0.2526.106 Safari/537
.36"
```

Matching of the first line for a multi-line log (with an IP address indicating the beginning of a line):

```
\d+\.\d+\.\d+\.\d+\s-\s. *
```

Regular expression used to extract log information:

```
(\S+) - - \[([^\]]*)] "(\S+) ([^"]+)" (\S+) (\S+) "([^"]+)" "([^"]+)"
```

Time conversion format:

```
%d/%b/%Y:%H:%M:%S
```

The following table describes extraction results of the sample log.

Key	Value
ip	**10.10.10.1**
time	**07/Jan/2016:21:06:39 +0800**
method	**GET**
url	**/wp-admin/js/password-strength-meter.min.js? ver=4.4 HTTP/1.0**

Key	Value
status	200
length	776
ref	http://wordpress.c4a1a0aecdb1943169555231dcc4adfb7.cn-hangzhou.alicontainer.com/wp-admin/install.php
user-agent	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_5) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/47.0.2526.106 Safari/537.36

18.8.3.8 Delimiter log

This topic describes the default format of delimiter logs to provide you with a reference for configuring Logtail to collect delimiter logs.

Log introduction

Delimiter logs use line breaks as boundaries. Each line represents a log. The fields in each log are connected by fixed delimiters, including single characters such as tabs, spaces, vertical bars (|), commas (,), and semicolons (;). If fields contain delimiters, you must enclose the fields in double quotation marks (").

Common delimiter logs include CSV and TSV formatted logs.

Log formats

A delimiter log is divided into several fields by delimiters, and supports two modes: single character and multiple characters.

- **Single character mode**

  In single character mode, log content is divided by single-character delimiters such as tabs (`\t`), spaces, vertical bars (`|`), commas (`,`), and semicolons (`;`).

  > **Note:**
  > **The double quotation mark (`"`) cannot be used as a delimiter. It is used as the quote of a single-character delimiter.**

  Single-character delimiters are often contained in log fields. To prevent log fields from being divided incorrectly, double quotation marks (`"`) are used as quotes to isolate log fields. If a double quotation mark (") is used as content

within a log field instead of as a quote, it must be escaped as `""`. You can use a double quotation mark (`"`) in a field boundary as a quote, or use a pair of double quotation marks (`""`) as field data. For situations that do not meet the format definition of delimiter logs, use modes such as the simple mode and full regex mode to parse fields.

- Double quotation mark (") used as a quote

  When a double quotation mark (`"`) is used as a quote, fields containing delimiters must be enclosed in a pair of quotes. Quotes must be located adjacent to delimiters. Modify the format if any characters such as spaces and tabs exist between quotes and delimiters.

  For example, when commas (`,`) function as delimiters, double quotation marks (") function as quotes, and the log format is `1997,Ford,E350,"ac, abs, moon",3000.00`, the log can be parsed into five fields: `1997`, `Ford`, `E350`, `ac, abs, moon`, and `3000.00`. Among these fields, `ac, abs, moon` enclosed in quotes is regarded as a complete field.
- Double quotation mark (") used as content within a log field

  If a double quotation mark (") is used as content within a log field instead of as a quote, it must be escaped as `""`. When this field is parsed, the double quotation marks (`""`) are restored to `"`.

  When commas function as delimiters and double quotation marks and commas are parts of a field, you must enclose the field with a pair of quotes and escape each double quotation mark (") into a pair of double quotation marks `""`. The log format after the processing is as follows:

  `1999,Chevy,"Venture ""Extended Edition, Very Large""",,5000.00`. The log can be parsed into five fields: `1999`, `Chevy`, `Venture "Extended Edition, Very Large"`, an empty field, and `5000.00`.

- Multiple characters mode

  In multiple characters mode, a delimiter can contain two or three characters, such as `||`, `&&&`, and `^_^`. In this mode, logs are parsed completely by matching delimiters and there is no need to enclose log fields within quotes.

  Note:

Make sure that the log field content does not contain complete delimiters. Otherwise, the log fields will be divided incorrectly.

For example, if the delimiter is set to `&&`, the log `1997&&Ford&&E350&&ac&abs&moon&&3000.00` will be parsed into five fields: `1997`, `Ford`, `E350`, `ac&abs&moon`, and `3000.00`.

Sample logs

- **A log with single-character delimiters**

```
05/May/2016:13:30:28,10.10.10.1,"POST /PutData? Category=YunOsAccou
ntOpLog&AccessKeyId=<yourAccessKeyId>&Date=Fri%2C%2028%20Jun%202013
%2006%3A53%3A30%20GMT&Topic=raw&Signature=<yourSignature> HTTP/1.1",
200,18204,aliyun-sdk-java
05/May/2016:13:31:23,10.10.10.2,"POST /PutData? Category=YunOsAccou
ntOpLog&AccessKeyId=<yourAccessKeyId>&Date=Fri%2C%2028%20Jun%202013
%2006%3A53%3A30%20GMT&Topic=raw&Signature=<yourSignature> HTTP/1.1",
401,23472,aliyun-sdk-java
```

- **A log with multi-character delimiters**

```
05/May/2016:13:30:28&&10.200.98.220&&POST /PutData? Category=
YunOsAccountOpLog&AccessKeyId=<yourAccessKeyId>&Date=Fri%2C%2028%
20Jun%202013%2006%3A53%3A30%20GMT&Topic=raw&Signature=<yourSignature
> HTTP/1.1&&200&&18204&&aliyun-sdk-java
05/May/2016:13:31:23&&10.200.98.221&&POST /PutData? Category=
YunOsAccountOpLog&AccessKeyId=<yourAccessKeyId>&Date=Fri%2C%2028%
20Jun%202013%2006%3A53%3A30%20GMT&Topic=raw&Signature=<yourSignature
> HTTP/1.1&&401&&23472&&aliyun-sdk-java
```

Configure Logtail to collect delimiter logs

For more information about how to configure Logtail to collect delimiter logs, see *Apache logs*. You can select the corresponding configuration based on your network deployment and the actual situation.

1. **Create a project and a Logstore. For more information about how to create a project and a Logstore, see *Create a project* and *Create a Logstore*.**
2. **On the Logstores page, click the icon in the Data Import Wizard column.**
3. **Select a data type.**

   Select Text File and click Next.

4. Configure a data source.
   a. Enter the configuration name and log path. Then, select Delimiter Mode as the log collection mode.
   b. Enter the sample log and select the delimiter.

      Select the appropriate delimiter based on the log format. Otherwise, parsing may fail.
   c. Specify keys in log extraction results.

      After you enter a sample log and select a delimiter, Log Service extracts fields of the log based on the delimiter and defines the fields as values. You must specify a key for each value.

      The preceding sample log uses commas (,) as delimiters and is divided into six fields. The keys for the fields are: time, ip, url, status, latency, and user-agent.
   d. Specify the log time.

      You can use the system time or the value of a log field such as the time field (05/May/2016:13:30:29 for example) as the log time. For more information about how to configure the date format, see *Configure a time format*.
   e. After the configuration is applied to the machine group, preview logs in the console to check whether logs are collected.

18.8.3.9 JSON logs

This topic describes the default format of JSON logs to provide you with a reference for configuring Logtail to collect JSON logs.

A JSON log can be written in two types of structures:

- Object: a collection of key-value pairs
- Array: an ordered list of values

Logtail supports JSON logs of the object type. Logtail automatically extracts the keys and values from the first layer of an object as the names and values of fields respectively. The field value can be an object, array, or basic types such as string or number.

Logtail does not support automatic parsing of non-object data such as JSON arrays. You need to use regular expressions to extract the fields or use the simple mode to collect logs by line.

Sample log

```
{"url": "POST /PutData? Category=YunOsAccountOpLog&AccessKeyId=<
yourAccessKeyId>&Date=Fri%2C%2028%20Jun%202013%2006%3A53%3A30%20GMT
&Topic=raw&Signature=<yourSignature> HTTP/1.1", "ip": "10.200.98.220
", "user-agent": "aliyun-sdk-java", "request": {"status": "200", "
latency": "18204"}, "time": "05/May/2016:13:30:28"}
{"url": "POST /PutData? Category=YunOsAccountOpLog&AccessKeyId=<
yourAccessKeyId>&Date=Fri%2C%2028%20Jun%202013%2006%3A53%3A30%20GMT
&Topic=raw&Signature=<yourSignature> HTTP/1.1", "ip": "10.200.98.210
", "user-agent": "aliyun-sdk-java", "request": {"status": "200", "
latency": "10204"}, "time": "05/May/2016:13:30:29"}
```

Configure Logtail to collect JSON logs

For more information about how to collect JSON logs by using Logtail, see *Apache logs*. Select configurations based on your network deployment and the actual situation.

1. **Create a project and a Logstore. For more information about how to create a project and a Logstore, see *Create a project* and *Create a Logstore*.**
2. **On the Logstores page, click the Data Import Wizard icon.**
3. **Select a data type.**

   Select Text File and click Next.
4. **Configure a data source.**

   a. **Enter the configuration name and Log Path, and set log collection mode to JSON mode.**

   b. **Determine whether to use the system time as the log time as needed. You can choose to enable or disable Use System Time.**

   - **Enable Use System Time**

     If this feature is enabled, the time Log Service collects a log is used as the log time instead of the time fields in the log.
   - **Disable Use System Time**

     If this feature is disabled, the time fields of the log are extracted and used as the log time.

     If you disable the Use System Time function, you must define the key of the extracted time field, and the time conversion format. For example, the `time` field (05/May/2016:13:30:29) in an object can be extracted as the log time.

For more information about how to set the date format, see *Configure a time format*.

5. **After the configuration is applied to the machine group, preview logs in the console to check whether logs are collected.**

18.8.3.10 ThinkPHP logs

ThinkPHP is a Web application development framework based on the PHP language.

ThinkPHP log format

Logs are printed in the following format in ThinkPHP:

```
<? php
Think\Log::record('D model class not found for method instantiation');
? >
```

Sample log

```
[ 2016-05-11T21:03:05+08:00 ] 10.10.10.1 /index.php
INFO: [ app_init ] --START--
INFO: Run Behavior\BuildLiteBehavior [ RunTime:0.000014s ]
INFO: [ app_init ] --END-- [ RunTime:0.000091s ]
INFO: [ app_begin ] --START--
INFO: Run Behavior\ReadHtmlCacheBehavior [ RunTime:0.000038s ]
INFO: [ app_begin ] --END-- [ RunTime:0.000076s ]
INFO: [ view_parse ] --START--
INFO: Run Behavior\ParseTemplateBehavior [ RunTime:0.000068s ]
INFO: [ view_parse ] --END-- [ RunTime:0.000104s ]
INFO: [ view_filter ] --START--
INFO: Run Behavior\WriteHtmlCacheBehavior [ RunTime:0.000032s ]
INFO: [ view_filter ] --END-- [ RunTime:0.000062s ]
INFO: [ app_end ] --START--
INFO: Run Behavior\ShowPageTraceBehavior [ RunTime:0.000032s ]
INFO: [ app_end ] --END-- [ RunTime:0.000070s ]
ERR: D model class not found for method instantiation
```

Configure Logtail to collect ThinkPHP logs

For more information about how to configure Logtail to collect Python logs, see *Apache logs*. **Select configurations based on your network deployment and the actual situation.**

The regular expression is automatically generated only based on the log sample and is not applicable to all logs. Therefore, you must adjust the regular expression.

ThinkPHP logs are multi-line logs that have varying modes. The following fields can be extracted from the ThinkPHP logs: time, source IP address of access, accessed

URL, and printed message. The message field contains multiple lines of information and has different modes. It can only be packaged into one field.

Parameters for configuring Logtail to collect ThinkPHP logs

Regular expression at the beginning of the line:

```
\[\s\d+-\d+-\w+:\d+:\d+\+\d+:\d+\s. *
```

Regular expression:

```
\[\s(\d+-\d+-\w+:\d+:\d+)[^:]+:\d+\s]\s+(\S+)\s(\S+)\s+(. *)
```

Time expression:

```
%Y-%m-%dT%H:%M:%S
```

18.8.3.11 Use Logstash to collect IIS logs

Before using Logstash to collect IIS logs, you must modify the configuration file to parse IIS log fields.

Sample log

View IIS log configurations, select W3C format (default field setting), and then save the modification.

```
2016-02-25 01:27:04 112.74.74.124 GET /goods/list/0/1.html - 80 - 66.
249.65.102 Mozilla/5.0+(compatible;+Googlebot/2.1;++http://www.google.
com/bot.html) 404 0 2 703
```

Collection configuration

```
input {
  file {
    type => "iis_log_1"
    path => ["C:/inetpub/logs/LogFiles/W3SVC1/*.log"]
    start_position => "beginning"
  }
}
filter {
  if [type] == "iis_log_1" {
  #ignore log comments
  if [message] =~ "^#" {
    drop {}
  }
  grok {
    # check that fields match your IIS log settings
    match => ["message", "%{TIMESTAMP_ISO8601:log_timestamp} %{
IPORHOST:site} %{WORD:method} %{URIPATH:page} %{NOTSPACE:querystring}
 %{NUMBER:port} %{NOTSPACE:username} %{IPORHOST:clienthost} %{NOTSPACE
:useragent} %{NUMBER:response} %{NUMBER:subresponse} %{NUMBER:scstatus
} %{NUMBER:time_taken}"]
  }
    date {
```

```
    match => [ "log_timestamp", "YYYY-MM-dd HH:mm:ss" ]
      timezone => "Etc/UTC"
  }
  useragent {
    source=> "useragent"
    prefix=> "browser"
  }
  mutate {
    remove_field => [ "log_timestamp"]
  }
  }
}
output {
  if [type] == "iis_log_1" {
  logservice {
        codec => "json"
        endpoint => "***"
        project => "***"
        logstore => "***"
        topic => ""
        source => ""
        access_key_id => "***"
        access_key_secret => "***"
        max_send_retry => 10
    }
    }
}
```

Note:

- **The configuration file must be encoded in UTF-8 without BOM. We recommend that you use Notepad++ to modify the file encoding format.**
- **The** *path* **field indicates a file path. When you specify a file path, you must use delimiters in UNIX format. Example:** *C:/test/multiline/*.log***. Otherwise, fuzzy matching cannot be performed.**
- **The** *type* **field must be modified in a unified manner and kept consistent in the file. If a machine has multiple Logstash configuration files, the type field in each configuration file must be unique. Otherwise, data cannot be processed correctly.**

Related plug-ins: *file* **and** *grok***.**

Restart Logstash to apply the configuration

Create a configuration file in the *conf* **directory. See** *Configure Logstash as a Windows service***. Restart Logstash to apply the configuration.**

18.8.3.12 Use Logstash to collect IIS logs

Before using Logstash to collect IIS logs, you must modify the configuration file to parse IIS log fields.

Sample log

View IIS log configurations, select W3C format (default field setting), and then save the modification.

```
2016-02-25 01:27:04 112.74.74.124 GET /goods/list/0/1.html - 80 - 66.
249.65.102 Mozilla/5.0+(compatible;+Googlebot/2.1;++http://www.google.
com/bot.html) 404 0 2 703
```

Collection configuration

```
input {
  file {
    type => "iis_log_1"
    path => ["C:/inetpub/logs/LogFiles/W3SVC1/*.log"]
    start_position => "beginning"
  }
}
filter {
  if [type] == "iis_log_1" {
  #ignore log comments
  if [message] =~ "^#" {
    drop {}
  }
  grok {
    # check that fields match your IIS log settings
    match => ["message", "%{TIMESTAMP_ISO8601:log_timestamp} %{
IPORHOST:site} %{WORD:method} %{URIPATH:page} %{NOTSPACE:querystring}
 %{NUMBER:port} %{NOTSPACE:username} %{IPORHOST:clienthost} %{NOTSPACE
:useragent} %{NUMBER:response} %{NUMBER:subresponse} %{NUMBER:scstatus
} %{NUMBER:time_taken}"]
  }
    date {
    match => [ "log_timestamp", "YYYY-MM-dd HH:mm:ss" ]
      timezone => "Etc/UTC"
  }
  useragent {
    source=> "useragent"
    prefix=> "browser"
  }
  mutate {
    remove_field => [ "log_timestamp"]
  }
  }
}
output {
  if [type] == "iis_log_1" {
  logservice {
        codec => "json"
        endpoint => "***"
        project => "***"
        logstore => "***"
        topic => ""
        source => ""
        access_key_id => "***"
```

```
        access_key_secret => "***"
        max_send_retry => 10
    }
    }
}
```

Note:

- **The configuration file must be encoded in UTF-8 without BOM. We recommend that you use Notepad++ to modify the file encoding format.**
- **The `path` field indicates a file path. When you specify a file path, you must use delimiters in UNIX format. Example: `C:/test/multiline/*.log`. Otherwise, fuzzy matching cannot be performed.**
- **The `type` field must be modified in a unified manner and kept consistent in the file. If a machine has multiple Logstash configuration files, the type field in each configuration file must be unique. Otherwise, data cannot be processed correctly.**

Related plug-ins: *file* and *grok*.

Restart Logstash to apply the configuration

Create a configuration file in the `conf` directory. See *Configure Logstash as a Windows service*. Restart Logstash to apply the configuration.

18.8.3.13 Use Logstash to collect other logs

Before using Logstash to collect logs, you can modify the configuration file to parse log fields.

Upload using the system time as the log time

- **Sample log**

```
2016-02-25 15:37:01 [main] INFO com.aliyun.sls.test_log4j - single
line log
2016-02-25 15:37:11 [main] ERROR com.aliyun.sls.test_log4j - catch
exception !
 java.lang.ArithmeticException: / by zero
    at com.aliyun.sls.test_log4j.divide(test_log4j.java:23) ~[bin
/:?]
    at com.aliyun.sls.test_log4j.main(test_log4j.java:13) [bin/:?]
2016-02-25 15:38:02 [main] INFO com.aliyun.sls.test_log4j - normal
log
```

- **Collection configuration**

```
input {
  file {
    type => "common_log_1"
```

```
    path => ["C:/test/multiline/*.log"]
    start_position => "beginning"
    codec => multiline {
      pattern => "^\d{4}-\d{2}-\d{2} \d{2}:\d{2}:\d{2}"
      negate => true
      auto_flush_interval => 3
      what => previous
    }
  }
}
output {
  if [type] == "common_log_1" {
  logservice {
        codec => "json"
        endpoint => "***"
        project => "***"
        logstore => "***"
        topic => ""
        source => ""
        access_key_id => "***"
        access_key_secret => "***"
        max_send_retry => 10
    }
    }
}
```

Note:

- **The configuration file must be encoded in UTF-8 without BOM. We recommend that you use Notepad++ to modify the file encoding format.**
- **The `path` field indicates a file path. When you specify a file path, you must use delimiters in UNIX format. For example, use `C:/test/multiline/*.log`. Otherwise, fuzzy matching cannot be performed.**
- **The `type` field must be modified in a unified manner and kept consistent in the file. If a machine has multiple Logstash configuration files, the `type` field in each configuration file must be unique. Otherwise, data cannot be processed correctly.**

Related plug-ins: *file*. For a single-line log file, you can remove the codec => multiline configuration.

- **Restart Logstash to apply the configuration**

  Create a configuration file in the `conf` directory. See *Configure Logstash as a Windows service*. Restart Logstash to apply the configuration.

18.9 Query and analysis

18.9.1 Indexing and querying

Log Service provides real-time LogSearch/Analytics capabilities for a large number of logs. When indexing is disabled, you can consume raw data in order based on shards, which is similar to ordered consumption of messages in Kafka. When indexing is enabled, you can perform statistical analysis and queries on log data in addition to ordered consumption.

Benefits

- **Real-time: Logs can be analyzed immediately after they are written.**
- **Fast:**
  - **Query: Billions of data records can be processed and queried within one second (with five conditions).**
  - **Analysis: Hundreds of millions of data records can be aggregated and analyzed within one second (aggregated with five dimensions and the GroupBy condition).**
- **Flexible: Query and analysis conditions can be changed as required and the results are returned in real time.**
- **Diverse ecosystem: In addition to the report, dashboard, and quick analysis feature provided in the console, LogSearch/Analytics also seamlessly integrates with Grafana, DataV, and Jaeger and supports RESTful APIs and JDBC APIs.**

Indexing

The indexing feature of Log Service can sort the values of log data in one or more columns, enabling you to quickly access the log data collected by Log Service. Before using LogSearch/Analytics, you must collect log data and *Configure an index* for the log data.

Log Service indexing consists of full-text indexing and key indexing.

- **Full-text indexing: Indexing is enabled for the full content of a log. The values of all the keys in the log are queried by default. The log can be queried if any of the keys matches the keyword.**

- **Key-value indexing: You can set different indexes for different keys. After setting key-value indexes for a log, you can query specific keys to narrow down the query scope.**

To use key-value indexing, you must specify the data type of each field. Log Service supports the text, numeric, and JSON types.

Terms

When LogSearch/Analytics (indexing) is disabled, you can consume raw data in order based on shards, which is similar to the ordered consumption of messages in Kafka. When indexing is enabled, you can perform statistical analysis and queries on log data in addition to ordered consumption.

Data types

You can configure the data type of each key in a log. A full-text index is a special key whose value is the entire log. Log Service supports the following data types.

Category	Type	Description	Example
Basic	*Text type*	The text type that supports matching by the combination of keywords and wildcards, and Chinese word segmentation.	`uri:"login*" method:"post"`
Basic	*Numeric type*	The numeric type that supports interval queries.	`status>200, status in [200, 500]`
Basic	*Numeric typeJSON type*	The type of floating-point numbers.	`price>28.95, t in [20.0, 37]`
Combination	*JSON type*	Indicates that the index is a JSON field that supports nested queries. The default field is text.	`level0.key>29.95 level0.key2:"action"`
Combination	*Text type*	Indicates that the full content of the log is queried as the text.	`error and "login fail"`

Syntax of LogSearch/Analytics

LogSearch/Analytics consists of Search and Analytics, which are separated with a vertical bar (|).

```
$Search |$Analytics
```

- **Search indicates the query condition, which can be generated by using keywords , wildcard queries, numeric values, intervals, and combinations of these data types. If it is left blank or if an asterisk (*) is used, all data is queried.**
- **Analytics is used to calculate and perform statistical analysis on query results or full data.**

> **Note:**
> **Both Search and Analytics are optional. If Search is empty, all the data in the specified period is not filtered and the results are counted directly. If Analytics is empty, the query results are returned and no statistical analysis is performed.**

Limits

If you query a large amount of log data (for example, the time span is long and the data volume exceeds 10 billion), you cannot query all the data in a single request . In this case, Log Service returns the existing data and notifies you that the query result is incomplete.

At the same time, the server caches the query results generated in the last 15 minutes. When the query result is partially cached, the server continues to scan log data that has not been cached. To reduce your workload of merging multiple query results, Log Service merges hit query results in the cache and new hit results of the current query and then returns them to you.

Therefore, Log Service enables you to obtain full results by calling this operation multiple times with the same parameters.

18.9.2 Real-time analysis

Log Service supports aggregate calculation. This feature combines queries with SQL calculation capacities to calculate the query result.

Syntax example:

```
status>200 |select avg(latency),max(latency) ,count(1) as c GROUP BY
method  ORDER BY c DESC  LIMIT 20
```

Basic syntax:

```
[search query] | [sql query]
```

A search condition and a calculation condition are separated by a vertical bar (|). This syntax indicates that a search query is used to filter the logs you need, and a SQL query is use to calculate these logs. The search query syntax is specific to Log Service. For more information, see *Query syntax*.

Prerequisites

To use the analysis feature, you must click Enable for the SQL-related fields in Search and Analysis Config.

- **If you do not enable the analysis feature, each shard calculates up to 10,000 rows of data but it takes a long time.**
- **After this feature is enabled, data analysis can be completed within seconds.**
- **Data generated after you enable this feature can be analyzed.**
- **No additional charges are incurred after this feature is enabled.**

Supported SQL syntax

Log Service supports the following SQL syntax. For more information, click the corresponding link.

- **Aggregate functions in SELECT statements:**
  - *General aggregate functions*
  - *Map functions*
  - *Approximate functions*
  - *Mathematical statistics functions*
  - *Mathematical calculation functions*
  - *String functions*
  - *Date and time functions*
  - *URL functions*
  - *Regular expression functions*
  - *JSON functions*
  - *Type conversion functions*
  - *Array functions*
  - *Binary string functions*
  - *Bitwise functions*
  - *Comparison functions and operators*
  - *Lambda functions*
  - *Logical functions*
  - *Geospatial functions*
- *GROUP BY syntax*
- *Window functions*
- *HAVING syntax*
- *ORDER BY syntax*
- *LIMIT syntax*
- *CASE WHEN syntax*
- *Column aliases*
- *Nested queries*

Syntax

The SQL syntax structure is as follows:

- The FROM and WHERE clause are not required in the SQL statements. The system uses the current Logstore for the FROM clause, and the search contition for the WHERE clause.
- The supported clauses include SELECT, GROUP BY, ORDER BY [ASC,DESC], LIMIT , and HAVING.
- Only the first 10 results are returned. To return more results, you must add limit n to the statement. For example, `* | select count(1) as c, ip group by ip order by c desc limit 100`.

Built-in fields

Log Service has built-in fields for statistical analysis. These built-in fields are automatically added when you configure a valid column.

Field name	Type	Definition
`__time__`	Bigint	The time when the log is generated.
`__source__`	Varchar	The source IP address of the log. Note: This field is source when you query data. The underscores (_) are added before and after source only in SQL.
`__topic__`	Varchar	The topic of the log.

Limits

1. The maximum number of concurrency for each project is 5.
2. A single varchar column has the maximum length of 512. The exceeded content will be truncated.
3. 100 rows of data are returned by default, and paging is not supported. If you want more data to be returned, use *LIMIT syntax*.

Example

Count the number of page views (PVs), unique visitors (UVs), and user requests with the top 10 latency per hour.

```
*|select date_trunc('hour',from_unixtime(__time__)) as time,
     count(1) as pv,
     approx_distinct(userid)  as uv,
     max_by(url,latency) as top_latency_url,
```

```
    max(latency,10) as top_10_latency
    group by  1
    order by time
```

18.9.3 Disable an index

You can disable an index when you do not need LogSearch/Analytics of Log Service.

Procedure

1. *Log on to the Log Service console*.
2. **Select a project and click the name of the project.**
3. **Select a Logstore and click Search in the LogSearch column.**
4. **Choose Index Attributes > Disable in the upper-right corner.**

18.9.4 Index data type

18.9.4.1 Overview

Log Service allows you to set full-text indexes or field indexes for the collected logs. If you set a full-text index for a log, the value is the entire log. If you set a field index for a log, you can set the data type of each key.

Data types

The following table describes the supported index types.

Category	Type	Description	Example
Basic	*Text type*	**The text type that supports matching by the combination of keywords and wildcards, and Chinese word segmentation.**	`uri:"login*" method:"post"`
Basic	*Numeric type*	**The numeric type that supports interval queries.**	`status>200, status in [200, 500]`
Basic	*Numeric typeJSON type*	**The type of floating-point numbers.**	`price>28.95, t in [20.0, 37]`
Combination	*JSON type*	**Indicates that the index is a JSON field that supports nested queries. The default field is text.**	`level0.key>29.95 level0.key2:"action"`

Category	Type	Description	Example
Combination	*Text type*	Indicates that the full content of the log is queried as the text.	`error and "login fail"`

Example

The following log includes time and other four keys.

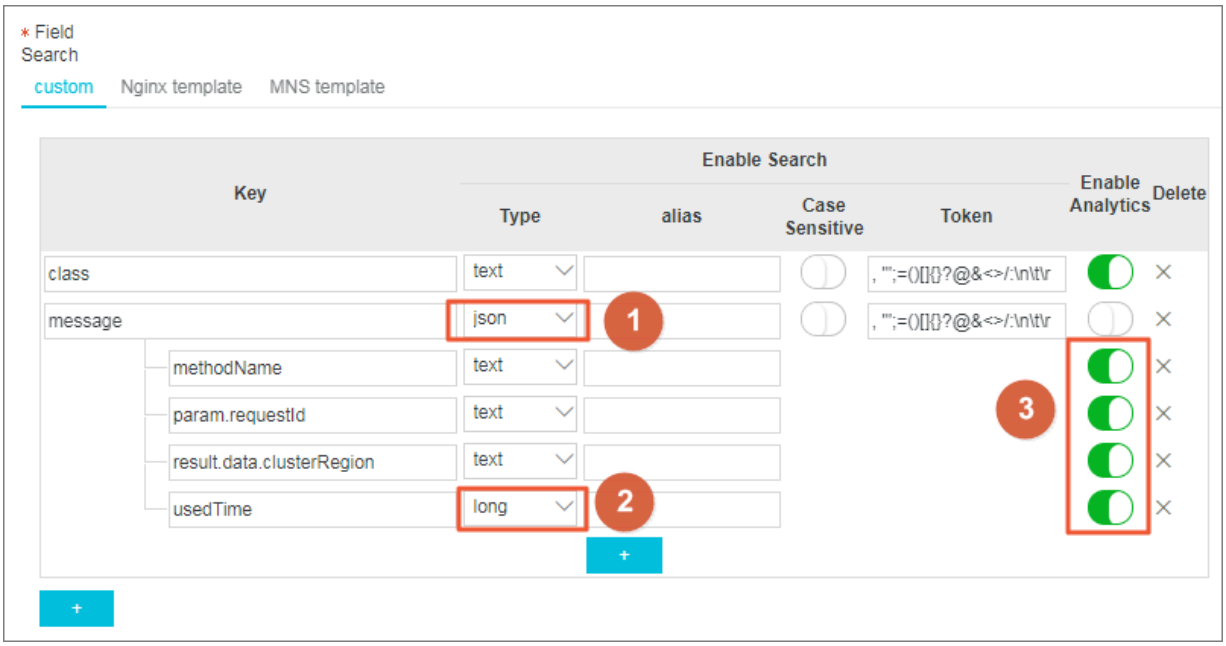
No.	Key	Type
0	time	-
1	class	Text
2	status	Long
3	latency	Double
4	message	JSON

```
0. time:2018-01-01 12:00:00
  1. class:central-log
  2. status:200
  3. latency:68.75
  4. message:
  {
      "methodName": "getProjectInfo",
      "success": true,
      "remoteAddress": "1.1.1.1:11111",
      "usedTime": 48,
      "param": {
              "projectName": "ali-log-test-project",
              "requestId": "d3f0c96a-51b0-4166-a850-f4175dde7323"
      },
      "result": {
          "message": "successful",
          "code": "200",
          "data": {
              "clusterRegion": "ap-southeast-1",
              "ProjectName": "ali-log-test-project",
              "CreateTime": "2017-06-08 20:22:41"
          },
          "success": true
      }
```

```
}
```

You can set indexes for a log as follows:

Figure 18-2: Index setting



In the preceding figure:

- **① indicates querying all the data of the string and bool types in JSON fields.**
- **② indicates querying data of the long type.**
- **③ indicates SQL analysis of configured fields.**

Example:

1. **Query of data of the string and bool types**
   - **You do not need to configure JSON fields.**
   - **JSON maps and arrays are automatically expanded. You can query multi-level nested fields by separating each level with a period (.).**

```
class : cental*
message.traceInfo.requestId : 92.137_1518139699935_5599
message.param.projectName : ali-log-test-project
message.success : true
```

2. **Query of data of the double and long types**

   JSON fields must be configured separately and cannot be contained in an array.

```
latency>40
```

```
message.usedTime > 40
```

3. **Combination query**

```
class : cental* and message.usedTime > 40 not message.param.
projectName:ali-log-test-project
```

18.9.4.2 Text type

Similar to search engines, text data is queried based on terms. Therefore, you must configure delimiters, case sensitivity, and Chinese word segmentation.

Configuration instructions

Case sensitivity

Specifies whether a raw log query is case-sensitive. For example, you want to query a raw log whose name is internalError.

- **false (case-insensitive) indicates that you can query a sample with either the keyword INTERNALERROR or internalerror.**
- **true (case-sensitive) indicates that you can query a sample log only with the keyword internalError.**

Delimiter

You can split the content of a raw log into several keywords by using a delimiter.

For example, you need to query the following log content:

```
/url/pic/abc.gif
```

- **If no delimiter is set, the entire string is considered as an individual word `/url/pic/abc.gif`. You can query this log only by using the entire string or performing fuzzy matching of `/url/pic/*`.**
- **If the delimiter is set to `/`, the raw log is split into three words: `url`, `pic`, and `abc.gif`. You can query this log by using any word or performing fuzzy matching of any word. For example, `url`, `abc.gif`, or `pi*`. You can also use `/url/pic/abc.gif`, which is resolved into `url AND pic AND abc.gif` when you query this log.**
- **If the delimiter is set to `/.`, the raw log is split into four words: `url`, `pic`, `abc`, and `gif`.**

Note:

You can extend query ranges by setting appropriate delimiters.

Full-text index

By default, all fields except the time field and keys of a log are considered as text data. You do not need to specify keys. For example, the following log is composed of the time field, the status field, the level field, and the message field:

```
[20180102 12:00:00] 200,error,some thing is
           error
          in this
           field
```

- **time:2018-01-02 12:00:00**
- **level:error**
- **status:200**
- **message:some thing is error in this field**

After you enable full-text index, the entire log is assembled into a piece of text data based on the key:value + "space" format. For example:

```
status:200 level:error message:"some thing is error in this field"
```

Note:

- **Prefixes are not required for full-text index. When you search the word error, logs with error in either the level field or the message field are queried.**
- **You must set delimiters for full-text index. When the delimiter is a space, status: 200 is considered as a phrase. When the delimiter is a colon (:), status and 200 are considered as two individual phrases.**
- **Numbers are processed as texts. For example, you can use 200 to query this log. The time field is not processed as text data.**
- **You can query the entire log if you enter a key such as status.**

18.9.4.3 Numeric type

When configuring indexes, you can configure a field as the numeric type and query the key by specifying a number range.

Configuration instructions

Supported types: `long` (long integers) and `double` (decimals). After configuring a field as the numeric type, you can only query the key by specifying a number range.

Query examples

To query the long-type key whose range is (1000 2000], you can use the following methods:

- **Query by specifying the numbers. Example:**

```
longKey > 1000 and longKey <= 2000
```

- **Query by specifying the number range. Example:**

```
longKey in (1000 2000]
```

For more information about syntax, see *Query syntax*.

18.9.4.4 JSON type

JSON is a combined data type that consists of text, boolean, number, array, and map .

Configuration instructions

- **Text type**

  Fields of the text type and bool type are automatically identified.

  For example, you can query the following keys in JSON format by using `jsonkey.key1:"text_value"` **and** `jsonkey.key2:true` **query conditions.**

```
jsonkey: {
   key1:text_value,
   key2:true,
   key3:3.14
```

```
}
```

- **Numeric type**

  You can query data of the double and long types in non-JSON arrays by setting a type and specifying a path.

  For example, the following query statement is used to query the jsonkey.key field of the double type:

  ```
  jsonkey.key3 > 3
  ```

- **Non-complete valid JSON**

  Log Service attempts to parse the valid content of the non-complete valid JSON data until the invalid content appears.

  Example:

  ```
  "json_string":
  {
       "key_1" :  "value_1",
       "key_map" :
        {
               "key_2" : "value_2",
               "key_3" : "valu
  ```

  The data following key_3 is truncated and lost. Log Service can correctly parse the `json_string.key_map.key_2` field and the content before this field.

Precautions

- **JSON object and JSON array types are not supported.**
- **Fields cannot be contained in JSON arrays.**
- **Fields of the bool type can be converted to fields of the text type.**

Query syntax

To query a specific key, you must add the parent path prefix in JSON. The query syntax for the text and numeric types is the same as other types. For more information, see *Query syntax*.

18.9.5 Query syntax and functions

18.9.5.1 Query syntax

This topic provides the syntax to express query conditions in Log Service, helping you to query logs more efficiently. You can query logs by calling the GetLogs and

GetHistograms operations in the Log Service API. You can also query logs by specifying query conditions on the query page in the Log Service console. This topic describes the syntax of query conditions.

Index types

Log Service enables you to create an index for a Logstore by using the following methods:

- **Full text index: queries a log entry as a whole without differentiating between keys and values.**
- **Key/value index: queries logs that contain the specified key. For example, you can use the keys *FILE:app* and *Type:action*. All the strings containing the specified keys are queried.**

Syntax keywords

LogSearch query conditions support the following keywords.

Name	Description
and	**Binary operator. The format is `query1 AND query2`, indicating the intersection of the query results of `query1` and `query2`. If no syntax keyword exists between queries, the `AND` operator is used to connect the queries by default.**
or	**Binary operator. The format is `query1 OR query2`, indicating the union of the query results of `query1` and `query2`.**
not	**Binary operator. The format is `query1 NOT query2`, indicating query results that match `query1` and do not match `query2`, which is equivalent to `query1-query2`. If only `NOT query1` exists, logs that do not contain `query1` are selected.**
(,)	**The left and right parentheses are used to merge multiple sub-queries into one query to increase the priority of the query in the parentheses.**
:	**The colon is used to query key-value pairs. `term1:term2` makes up a key-value pair. If the key or value contains reserved characters such as spaces and colons (`:`), use double quotation marks (`""`) to enclose the entire key or value.**

Name	Description
"	The double quotation mark is used to convert a keyword into a common query character. All the terms enclosed in double quotation marks are queried and are not considered as syntax keywords. Alternatively, all the terms enclosed in double quotation marks are regarded as a whole.
\	Escape character. It is used to escape quotation marks. The escaped quotation marks indicate the symbols themselves and are not considered as escape characters such as `"\""`.
|	Pipeline operator, indicating that more computations are performed based on the previous computation. Example: query1 | timeslice 1h | count.
timeslice	Timeslice operator, indicating that time-sliced data is calculated as a whole. You can use timeslice 1h to indicate a one-hour time slice, timeslice 1m to indicate a one-minute time slice, and timeslice 1s to indicate a one-second time slice. For example, query1 | timeslice 1h | count indicates that query1 is queried as the query condition, and the total number of results in one hour is returned.
count	Count operator, indicating the number of logs.
*	Wildcard character. It is used to perform fuzzy matching and can replace none or multiple characters. For example, if `que*` is used in a query, all words that start with `que` are returned.  **Note:** A maximum of 100 logs with words that match the keyword are returned.
?	The wildcard character. It is used to perform fuzzy matching and can replace only one character. For example, if `qu? ry` is used in a query, all words starting with `qu`, ending with `ry`, and with a character in between are returned.
`__topic__`	Queries data of a certain topic. You can query the data of zero or multiple topics in the query. Example: `__topic__:mytopicname`.

Name	Description
`__tag__`	**Queries a tag value of a tag key. Example:** `__tag__:tagkey:tagvalue`.
source	**Queries data of an IP address. Example:** `source:127.0.0.1`.
>	**Queries the logs with the value of a field greater than a specific number. Example:** `latency > 100`.
>=	**Queries the logs with the value of a field greater than or equal to a specific number. Example:** `latency >= 100`.
<	**Queries the logs with the value of a field smaller than a specific number. Example:** `latency < 100`.
<=	**Queries the logs with the value of a field smaller than or equal to a specific number. Example:** `latency <= 100`.
=	**Queries the logs with the value of a field equal to a specific number. Example:** `latency = 100`.
in	**Queries the logs with a field falling within a specific range. Brackets ([]) are used to indicate closed intervals and parentheses (()) are used to indicate open intervals. Two numbers are enclosed in brackets or parentheses and separated by multiple spaces. Example:** `latency in [100 200]` **or** `latency in (100 200]`.

Note:

- **Syntax keywords are case-insensitive.**
- **Priorities of syntax keywords are sorted in descending order as follows:** `: > " > ( ) > and not > or`.
- **Log Service reserves the right to use the following keywords:** `sort asc desc group by avg sum min max limit`. **If you need to use these keywords, enclose them with double quotation marks ("").**
- **If you want to use full-text index and key-value index at the same time, you must configure the same delimiter for these methods. Otherwise, no query results will be returned for full-text index.**
- **To perform a numeric query, set the data type of the queried column to double or long. If no data type is set or the syntax used for the numeric range query is**

incorrect, Log Service translates the query condition into a full text index, which may lead to an unexpected result.

- **If you change the data type of a column from text to numeric, only the = query is supported for the data prior to this change.**

Query examples

1. **Logs that contain a and b at the same time:** `a AND b` **or** `a b`**.**
2. **Logs that contain a or b:** `a OR b`**.**
3. **Logs that contain a but do not contain b:** `a NOT b`**.**
4. **Logs that do not contain a:** `NOT a`**.**
5. **Logs that contain a and b, but do not contain c:** `a AND b NOT c`**.**
6. **Logs that contain a or b and must contain c:** `(a OR b ) AND c`**.**
7. **Logs that contain a or b, but do not contain c:** `(a OR b ) NOT c`**.**
8. **Logs that contain a and b and may contain c:** `a AND b OR c`**.**
9. **Logs with the FILE field containing apsara:** `FILE:apsara`**.**
10. **Logs with the FILE field containing apsara and shennong:** `FILE:"apsara shennong"`**,** `FILE:apsara FILE: shennong`**, or** `FILE:apsara AND FILE:shennong`**.**
11. **Logs containing and:** `and`**.**
12. **Logs with the FILE field containing apsara or shenong:** `FILE:apsara OR FILE: shennong`**.**
13. **Logs with the file info field containing apsara:** `"file info":apsara`**.**
14. **Logs that contain quotation marks:** `\"`**.**
15. **Logs starting with shen:** `shen*`**.**
16. **Logs starting with shen in the FILE field:** `FILE:shen*`**.**
17. **Logs starting with shen, ending with ong, and with only one character in the middle:** `shen? ong`**.**
18. **Logs starting with shen and aps:** `shen* and aps*`**.**
19. **The distribution of logs starting with shen, with a time slice of 20 minutes:** `shen *| timeslice 20m | count`**.**
20. **All the data from topic 1 and topic 2:** `__topic__:topic1 or __topic__ : topic2` **.**
21. **All the data of tagvalue 2 in tagkey 1:** `__tag__ : tagkey1 : tagvalue2`**.**

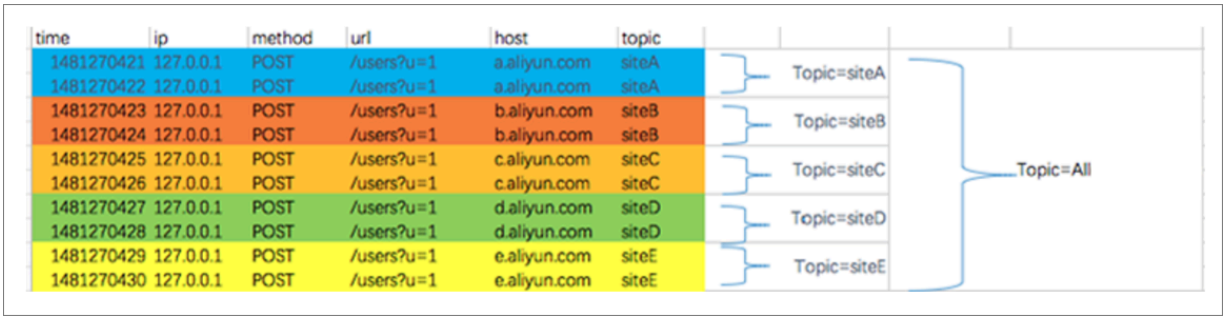
22. A query for all the data with a latency greater than or equal to 100 and less than 200 can be written in either of the following ways: `latency >=100 and latency < 200` or `latency in [100 200)`.
23. A query for all the requests with a latency greater than 100 must be written in the following way: `latency > 100`.
24. Logs that do not contain crawlers and do not contain opx in http_referer: `not spider not bot not http_referer:opx`.
25. Logs with the empty cdnIP field: `not cdnIP:""`.
26. Logs without the cdnIP field: `not cdnIP:*`.
27. Logs with the cdnIP field: `cdnIP:*`.

Specified or cross-topic query

Each Logstore can be divided into one or more subspaces based on the topic. You can specify topics to limit the query range when querying logs. This can help increase query performance. Therefore, we recommend that you use topics to divide a Logstore if you have a secondary classification requirement for the Logstore.

When you specify one or more topics, you can only query the data from topics that meet the query conditions. However, if no topic is specified, the data from all topics is queried by default.

For example, topics are used to classify logs with different domain names.



Topic query syntax

- The data from all topics can be queried. The data from all topics is queried if no topic is specified in the query syntax and parameters.
- A topic can be queried in a query. The query syntax is `__topic__:topicName`. The previous mode of specifying a topic in the URL parameter is still supported.

- **Multiple topics can be queried. For example, `__topic__:topic1 or __topic__:topic2` indicates that the union of data from topic 1 and topic 2 is queried.**

Fuzzy matching

Log Service supports fuzzy matching. You can specify a keyword no more than 64 characters, and use an asterisk (`*`) or a question mark (`?`) at the end or in the middle of the keyword. 100 words that match the keyword will be queried, and all logs that contain the 100 words and match the query conditions will be returned.

Note:

- **You must specify a prefix when querying logs. An asterisk (`*`) or a question mark (`?`) cannot be used at the beginning of a keyword.**
- **The more precise the keyword is, the more accurate result you will get.**
- **You cannot use fuzzy matching to search a keyword that exceeds 64 characters. We recommend you specify a keyword no less than 64 characters.**

18.9.5.2 Context query

When you expand a log file, you can see several log entries. Each log entry records an event and does not exist independently. Several consecutive log entries can help you review the process of an event in sequence.

Prerequisites

- *You can use Logtail to collect data* **and upload data to the Logstore. No other configurations are required except for creating machine groups and log collection configurations. You can also use producer SDKs to upload data, such as Producer Library, Log4j, logback, and C-producer library.**
- **Enable indexing.**

Note:

The context query is not applicable to syslogs.

Log context query specifies the log source (machine + files) and a log entry. It also queries several log entries before and after the log entry in the original log file. It provides an effective way to troubleshoot problems in DevOps scenarios.

The Log Service console provides a page for query, where you can view the context information of a specified log in the original file. This is similar to paging up and

down in the original log file. You can quickly locate error messages during business troubleshooting by viewing the context information of a specified log.

Scenarios

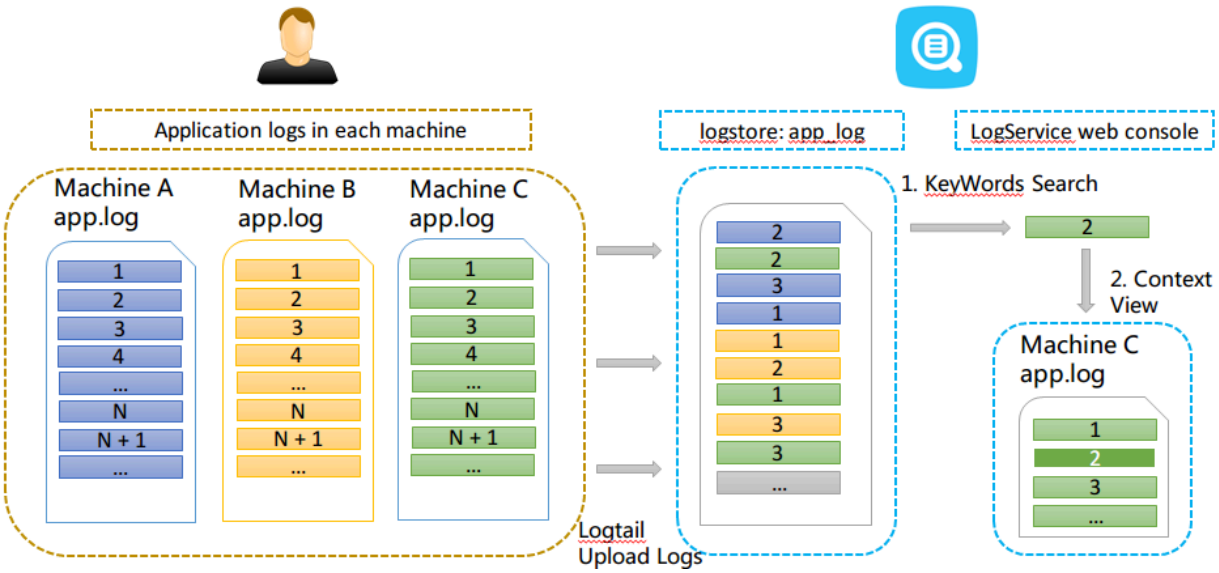
For example, an O2O take-out website will record the transaction path of an order in a program log on the server:

User logon > Browse products > Click items > Add to the shopping cart > Place an order > Pay for the order > Deduct payment > Generate an order

If you fail to place an order, the Operation & Maintenance (O&M) personnel must quickly identify the cause. Conventionally, the administrator grants the machine logon permission to related personnel. Then, the investigator logs on to each machine where applications are deployed, and uses the order ID as the keyword to search application logs, attempting to identify the cause.

In Log Service, you can perform the following steps to troubleshoot problems:

1. Install Logtail on the server, and add machine groups and log collection configurations in the console. Logtail then starts to upload incremental logs.
2. On the log query page in the Log Service console, specify the time range and find the error log based on the order ID.
3. Based on the error log, page up until all other related logs are found. For example, your credit card payment fails to be collected.



Benefits

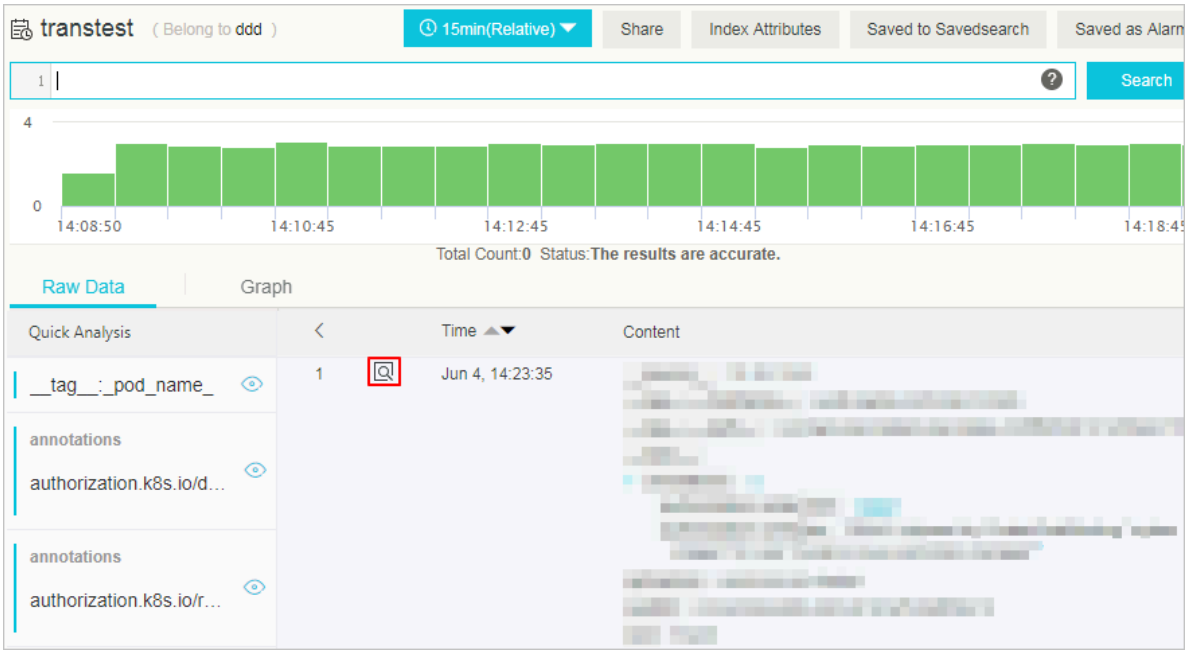
- There is no intrusion into applications and no need to modify the formats of log files.
- You can view the log context information of any machine or file in the Log Service console, without logging on to each machine to view log files.
- You can specify the time range based on the time of occurrence of the events to quickly locate suspicious logs and then perform context query in the Log Service console. This significantly improves troubleshooting efficiency.
- You do not need to worry about data loss caused by insufficient server storage or log file rotation, and can view historical data in the Log Service console at any time.

Procedure

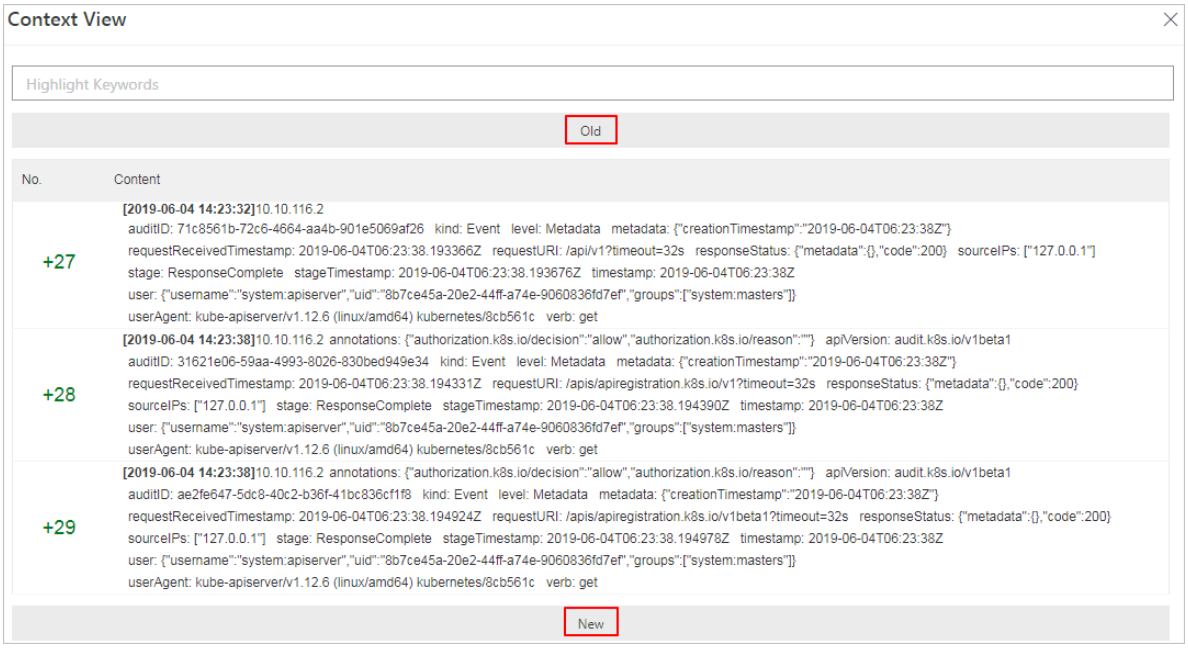
1. *Log on to the Log Service console*.
2. Select the desired project and click the project name.
3. On the Logstores page, select the desired Logstore, and click Search in the LogSearch column to go to the query page.

   If the Context View button appears on the left-side of a log on the query results page, the log supports context query.
4. Enter your query and analysis statement, select a time range, and click Search.

   If the Context View button appears on the left-side of a log on the query results page, the log supports context query.



5. Select a log and click Context View. On the page that appears on the right, view the context log of the target log.
6. Scroll up and down to view the context information of the selected log. To view more context logs, click Earlier or Later.



18.9.5.3 Other features

In addition to the statement-based queries, LogSearch/Analytics of Log Service provides the following extended features for query optimization.

Raw logs

After the index is enabled, enter the keywords in the search box and select the search time range. Then, click Search to view the histogram of the number of logs, the raw logs, and the statistical graph.

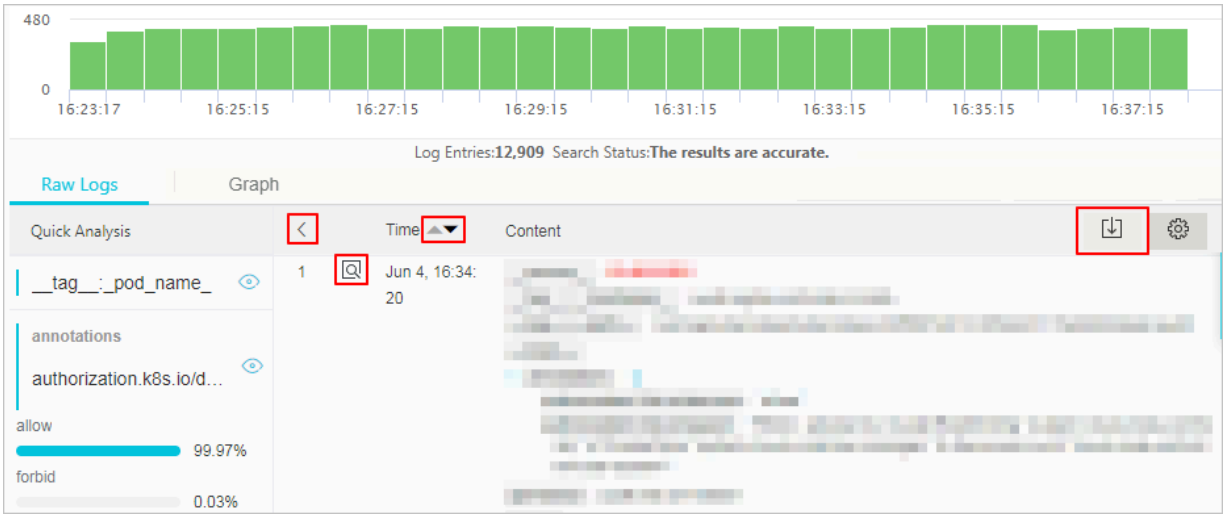
The histogram of the number of logs displays the time-based distribution of log search hit counts. On the histogram, you can view the changes in the number of logs over a certain period of time. You can click the rectangular to view the information about the log hits within the specified time. This can help refine the display of log search results.

On the Raw Logs tab, you can view the hit logs in chronological order.

- **Click the triangle symbo beside Time to switch between chronological order and reverse chronological order.**
- **Click the value keyword in the log content to view all logs that contain this keyword.**

- **Click the Download icon in the upper-right corner of the Raw Logs tab to download the query results in CSV format. Click Column Settings to add fields as displayed columns in the displayed results of raw logs. This can help you view the target field content of each raw log in the new columns in a more intuitive way.**
- **Click Context to view 15 log entries before and after the current log entry. For more information, see** *Context query*.

> **Note:**
> **The context query feature supports only the data uploaded with Logtail.**



Context query

The Log Service console provides a page for query, where you can view the context information of a specified log in the original file. This is similar to paging up and down in the original log file. You can quickly locate error messages during business troubleshooting by viewing the context information of a specified log. For more information, see *Context query*.

Quick analysis

The quick analysis feature of Log Service supports one-click interactive queries, helping you quickly analyze the distribution of a field during a period of time and reducing the cost of indexing key data. For more information, see *Quick analysis*.

Saved search

On the query page, click Saved Search in the upper-right corner to save your current query action as a saved search. You can initiate the query action again on the Saved Search tab without entering the query statement manually.

If you have added the saved search to Tag, you can directly access it in tags.

Configuring alerts

Log Service enables you to configure alerts based on your LogSearch results. You can configure alert rules so that specific alert content can be sent to you in the form of in-site notifications or DingTalk messages.

The basic process is as follows:

1. **Configure saved searches.**
2. **Configure alert rules.**
3. **Configure notification methods.**
4. **View alert records.**

For more information, see *Configure an alert*.

18.9.5.4 Quick analysis

The quick analysis feature of Log Service supports one-click interactive queries, helping you quickly analyze the distribution of a field over a period of time and reducing the cost of indexing key data.

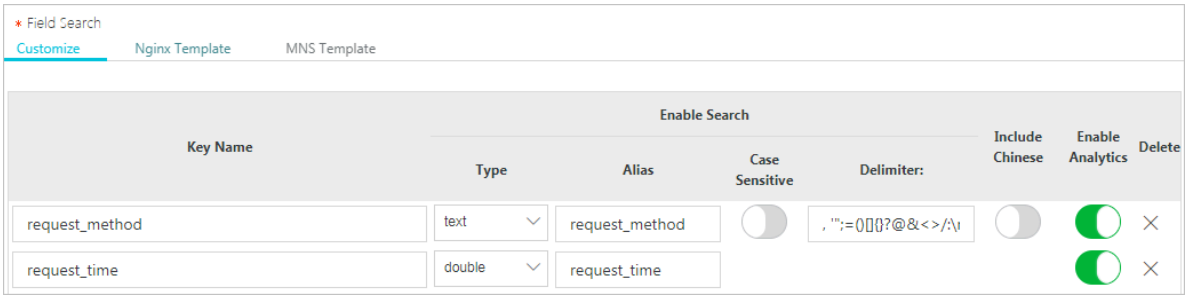
Features

- **Groups the first 10,000 pieces of data of the `text` type and provides statistics about the top 10 by groups.**
- **Supports quick generation of `approx_distinct` query statements for `text` fields.**
- **Supports histogram statistics about the approximate distribution of `long` or `double` fields.**
- **Supports quick search for the maximum, minimum, average, or sum of `long` or `double` fields.**
- **Generates a query statement based on quick search analysis.**

You must specify the field query properties before using quick analysis.

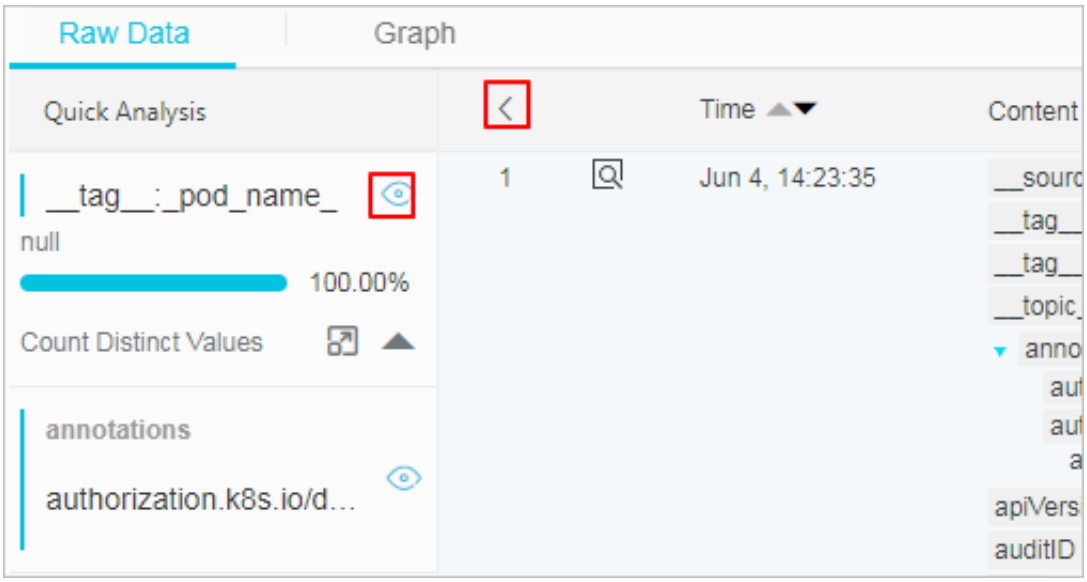
1. **Before specifying field query attributes, you must enable indexing for query and analysis.**
2. **Set `key` in a log as the field name and set the type, alias, and delimiter.**

   If an access log contains the `request_method` field and the `request_time` field, you can configure the settings as follows.



Instructions

After setting the specified field query, you can go to the query page, click the Raw Logs tab, and view the fields in the left-side Quick Analysis column. Click the button above the serial number, you can fold the page. Click Eye button to perform quick analysis based on the Current Temporal Interval and Current $Search conditions.



Text type

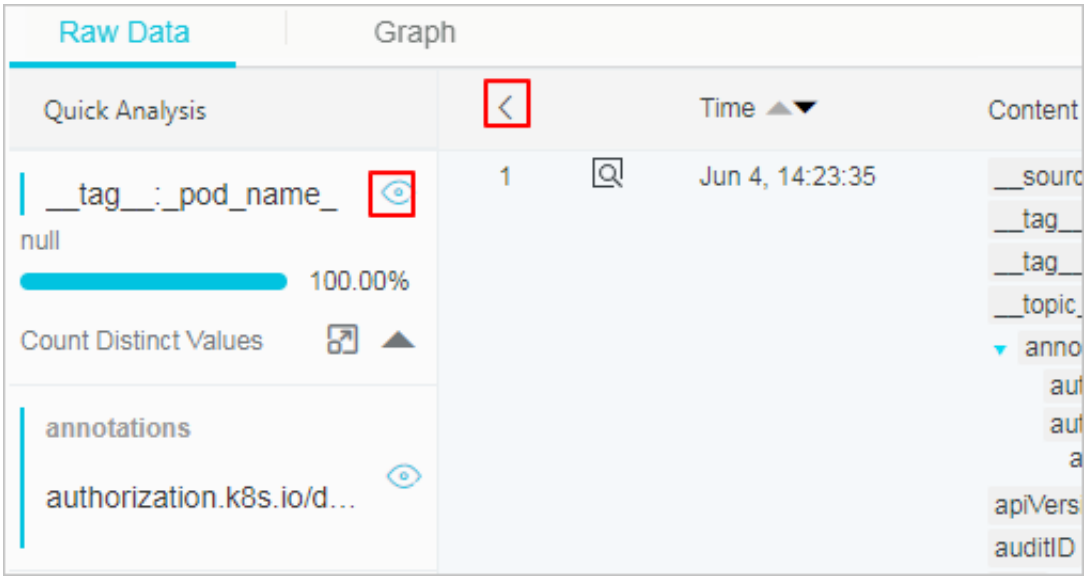
- **Statistics about text fields by group**

  Click the Eye button on the right of the field to quickly group the first 10,000 pieces of data of the text type and return the ratio of the top 10 groups.

  The following query statement is used:

```
$Search | select count(1) as pv , "${keyName}"  from ( select "${
keyName}" from log limit 10000) group by "${keyName}" order by pv
desc limit 10
```

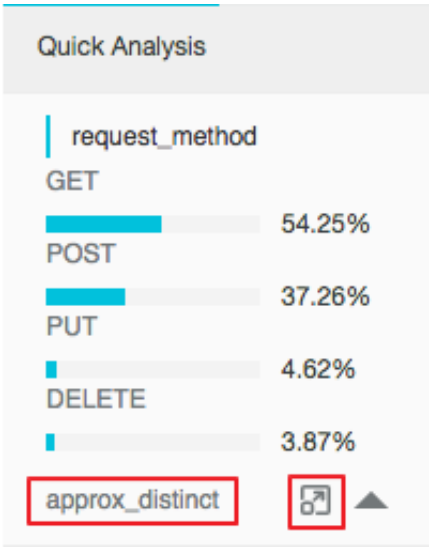
  When you query `request_method`, you can obtain the following result based on grouping statistics, where `GET` requests account for the majority of request methods.



- **Check the number of unique entries of the field**

  Under the target fields in the Quick Analysis column, click approx_distinct to check the number of unique entries for `${keyName}`.

  When you query `request_method`, you can obtain the following result based on grouping statistics, where `GET` requests account for the majority of request methods.



- **Extend the query statement of grouping statistics to the search box**

  Click the button to the right of approx_distinct to extend the query statement of grouping statistics to the search box for further operations.

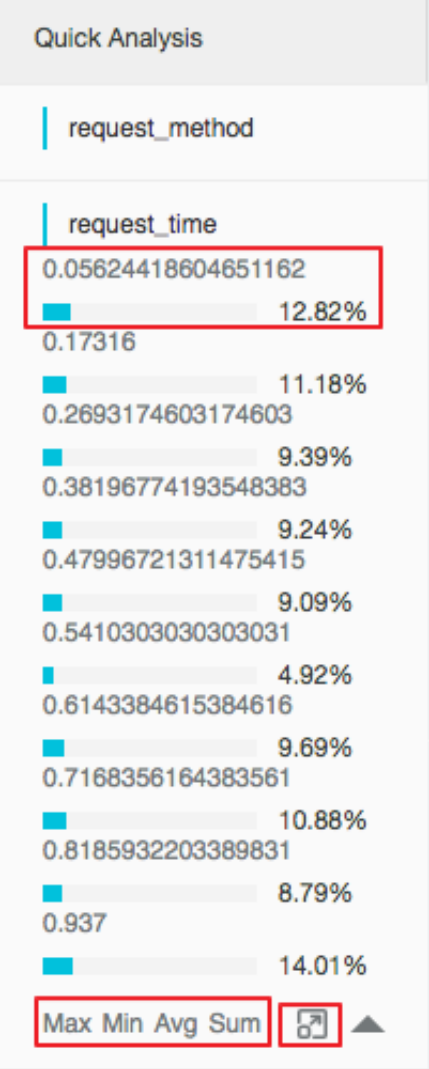
Long and double types

- **Approximate distribution using histograms**

  Grouping statistics is of little significance for the `long and double` types, which have multiple type values. Therefore, you can split the values into 10 buckets to obtain the approximate distribution. The following query statement is used:

  ```
  $Search | select numeric_histogram(10, ${keyName})
  ```

  When you query `request_time`, you can obtain the following approximate distribution results. From the result, you can see that the request time is mostly distributed around 0.059.



- **Quick analysis of the `Max`, `Min`, `Avg`, and `Sum` statements**

  Click `Max`, `Min`, `Avg`, and `Sum` under a field to quickly search for the maximum value, minimum value, average value, and sum of `Max`.

- **Extend the query statement of grouping statistics to the search box**

  Click the button to the right of `Sum` to extend the query statement of the histogram statistics for the approximate distribution to the search box for further operations.

18.9.5.5 Saved search

Saved search is a one-click query and analysis function provided by Log Service.

Prerequisites

You have enabled and configured Index.

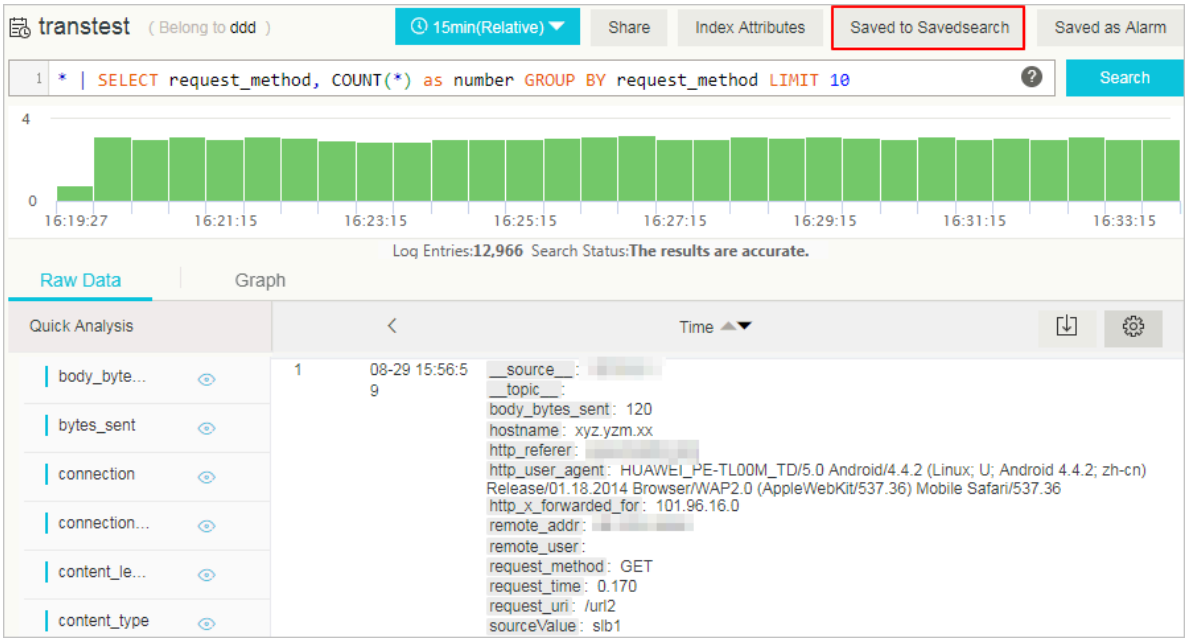
Context

If you need to frequently view the results of a query and analysis statement, save the statement as a saved search. In later searches, you only need to click the name of the saved search on the left side of the search page, eliminating the need to manually enter the statement again. You can also use saved searches in alert rules . Log Service executes a saved search periodically and sends an alert notification when the search result meets the preset condition of the statement.

Procedure

1. *Log on to the Log Service console*.
2. **Click a project name.**
3. **Click Search in the LogSearch column on the Logstores page.**
4. **Enter your query and analysis statement, set the time range, and click Search & Analysis.**

5. Click Save Search in the upper-right corner of the page.



6. Configure saved search attributes.

   a) Set Operation Type.

   - Save Search: saves the current query statement as a new saved search.
   - Modify Saved Search: modifies an existing saved search.

   b) Set Saved Search Name.

   - The name can only contain lowercase letters, digits, hyphens (-), and underscores (_).
   - The name must start and end with a lowercase letter or digit.
   - This name must be 3 to 63 characters in length.

   c) Confirm Logstores, Topic, and Query.

   If Logstores and Topic do not meet your requirements, return to the search page to access the target Logstore and enter your query statement, and then click Save Search again.

7. Click OK.

18.9.6 Analysis syntax and functions

18.9.6.1 General aggregate functions

The query and analysis feature of Log Service supports general aggregate functions.

The query and analysis feature of Log Service allows you to use general aggregate functions for log analysis. The following table describes the supported general aggregate functions.

Function	Description	Example
`arbitrary(x)`	**Returns an arbitrary value of x.**	`latency > 100 | select arbitrary(method)`
`avg(x)`	**Returns the average (arithmetic mean) of all input values.**	`latency > 100 | select avg(latency)`
`checksum(x)`	**Returns a Base64-encoded checksum of the given values.**	`latency > 100 | select checksum(method)`
`count(*)`	**Returns the number of input rows.**	**N/A**
`count(x)`	**Returns the number of non-null input values.**	`latency > 100 | count( method)`
`count_if(x)`	**Returns the number of TRUE input values.**	`latency > 100 | count( url like '%abc')`
`geometric_mean(x)`	**Returns the geometric mean of all input values.**	`latency > 100 | select geometric_mean(latency)`
`max_by(x,y)`	**Returns the value of x associated with the maximum value of y over all input values.**	`latency>100 | select max_by(method,latency)`
`max_by(x,y,n)`	**Returns n values of x associated with the n largest of all input values of y.**	`latency > 100 | select max_by(method,latency ,3)`
`min_by(x,y)`	**Returns the value of x associated with the minimum value of y over all input values.**	`* | select min_by(x,y)`

Function	Description	Example
`min_by(x,y,n)`	**Returns n values of x associated with the n smallest of all input values of y.**	`* | select max_by( method,latency,3)`
`max(x)`	**Returns the maximum value of all input values.**	`latency > 100| select max(inflow)`
`min(x)`	**Returns the minimum value of all input values.**	`latency > 100| select min(inflow)`
`sum(x)`	**Returns the sum of all input values.**	`latency > 10 | select sum(inflow)`
`bitwise_and_agg(x)`	**Returns the bitwise AND of all input values in two's complement representation.**	**N/A**
`bitwise_or_agg(x)`	**Returns the bitwise OR of all input values in two's complement representation.**	**N/A**

18.9.6.2 Map functions

The query and analysis feature of Log Service allows you to use map functions for log analysis.

The following table describes the supported map functions.

Function	Description	Example
Subscript operator []	**Retrieves the value corresponding to a given key from a map.**	**N/A**
histogram(x)	**Returns a map containing the count of the number of times each input value occurs. The syntax is equivalent to** `select count group by x` **.**	`latency > 10 | histogram(status)`, **which is equivalent to** `latency > 10 | select count(1) group by status`**.**

Function	Description	Example
map_agg(key,value)	Returns a map created from the input key/value pairs, and shows the random value of each key.	`latency > 100 | select map_agg(method,latency)`
multimap_agg(key,value)	Returns a multimap created from the input key /value pairs, and shows all the values of each key.	`latency > 100 | select multimap_agg(method, latency)`
cardinality(x) → bigint	Returns the cardinality (size) of the map x.	N/A
element_at(map<K, V>, key) → V	Returns the value for the given key.	N/A
map() → map<unknown, unknown>	Returns an empty map.	N/A
map(array<K>, array<V>) → map<K,V>	Returns a map created using the given key/value arrays.	`SELECT map(ARRAY[1,3], ARRAY[2,4]); – {1 -> 2 , 3 -> 4}`
map_from_entries(array<row<K, V>>) → map<K,V>	Returns a map created from the given array of entries.	`SELECT map_from_entries (ARRAY[(1, ‘x’), (2, ‘ y’)]); – {1 -> ‘x’, 2 - > ‘y’}`
map_entries(map<K, V>) → array<row<K,V>>	Returns an array of all entries in the given map.	`SELECT map_entries(MAP (ARRAY[1, 2], ARRAY[‘x ’, ‘y’])); – [ROW(1, ‘x ’), ROW(2, ‘y’)]`
map_concat(map1<K, V>, map2<K, V>, ⋯, mapN<K , V>) → map<K,V>	Returns the union of all the given maps. If a key is found in multiple given maps, the value of the key in the resulting map comes from the last one of those maps.	N/A
map_filter(map<K, V>, function) → map<K,V>	Refer to the lambda map_filter function.	N/A
transform_keys(map<K1 , V>, function) → MAP<K2 ,V>	Refer to the lambda transform_keys function.	N/A

Function	Description	Example
transform_values(map<K, V1>, function) → MAP<K,V2>	Refer to the lambda transform_values function.	N/A
map_keys(x<K, V>) → array<K>	Returns an array of all the keys in the map x.	N/A
map_values(x<K, V>) → array<V>	Returns an array of all the values in the map x.	N/A
map_zip_with(map<K, V1>, map<K, V2>, function<K, V1, V2, V3>) → map<K,V3>	Refer to the lambda map_zip_with function.	N/A

18.9.6.3 Approximate functions

The query and analysis feature of Log Service allows you to use approximate functions for log analysis.

The following table describes the supported approximate functions.

Function	Description	Example
approx_distinct(x)	Returns the approximate number of distinct input values.	N/A
approx_percentile(x, percentage)	Returns the approximate percentile for all input values of x at the given percentage.	approx_percentile(x,0.5)
approx_percentile(x, percentages)	Returns the approximate percentile for all input values of x at each of the specified percentages.	approx_percentile(x, array(0.1,0.2))

Function	Description	Example
`numeric_histogram( buckets, value)`	**Computes an approximate histogram with up to specified number of buckets for all values. The key and item count for each bucket are returned. This function is equivalent to `select count group by`.**	`method:POST | select numeric_histogram(10, latency)`

18.9.6.4 Mathematical statistics functions

The query and analysis feature of Log Service allows you to use mathematical statistics functions for log analysis.

Function	Description	Example
`corr(y, x)`	**Returns the correlation coefficient of input values. The result ranges from 0 to 1.**	`latency>100| select corr(latency,request_size)`
`covar_pop(y, x)`	**Returns the population covariance of input values.**	`latency>100| select covar_pop(request_size, latency)`
`covar_samp(y, x)`	**Returns the sample covariance of input values.**	`latency>100| select covar_samp(request_size,latency)`
`regr_intercept(y, x)`	**Returns the linear regression intercept of input values. y is the dependent value. x is the independent value.**	`latency>100| select regr_intercept( request_size,latency)`
`regr_slope(y,x)`	**Returns the linear regression slope of input values. y is the dependent value. x is the independent value.**	`latency>100| select regr_slope(request_size,latency)`
`stddev(x)` **or** `stddev_samp (x)`	**Returns the sample standard deviation of all input values.**	`latency>100| select stddev(latency)`

Function	Description	Example
`stddev_pop(x)`	**Returns the population standard deviation of all input values.**	`latency>100| select stddev_pop(latency)`
`variance(x)` **or** `var_samp(x)`	**Returns the sample variance of all input values.**	`latency>100| select variance(latency)`
`var_pop(x)`	**Returns the population variance of all input values.**	`latency>100| select variance(latency)`

18.9.6.5 Mathematical calculation functions

The query and analysis feature of Log Service allows you to use mathematical calculation functions for log analysis.

By using mathematical calculation functions with query statements, you can perform mathematical calculation to the log query results.

Mathematical operators

Mathematical operators include the plus sign (+), minus sign (-), multiplication sign (*), division sign (/), and percent sign (%). These operators can be used in the SELECT clause.

Example:

```
*|select avg(latency)/100 , sum(latency)/count(1)
```

Description of mathematical calculation functions

Log Service supports the following mathematical functions.

Function	Description
`abs(x)`	**Returns the absolute value of x.**
`cbrt(x)`	**Returns the cube root of x.**
`ceiling(x)`	**Returns x rounded up to the nearest integer.**
`cosine_similarity(x,y)`	**Returns the cosine similarity between the sparse vectors x and y.**
`degrees`	**Converts the angle in radians to degrees.**
`e()`	**Returns Euler’s number.**

Function	Description
`exp(x)`	Returns Euler’s number raised to the power of x.
`floor(x)`	Returns x rounded down to the nearest integer.
`from_base(string,radix)`	Returns the value of string interpreted as a radix number.
`ln(x)`	Returns the natural logarithm of x.
`log2(x)`	Returns the base 2 logarithm of x.
`log10(x)`	Returns the base 10 logarithm of x.
`log(x,b)`	Returns the base b logarithm of x.
`pi()`	Returns the constant Pi.
`pow(x,b)`	Returns x raised to the power of b.
`radians(x)`	Converts angle x in degrees to radians.
`rand()`	Returns a random number.
`random(0,n)`	Returns a random number between 0 and n (exclusive).
`round(x)`	Returns x rounded to the nearest integer .
`round(x, y)`	Returns x rounded to y decimal places. For example, round(1.012345,2) = 1.01.
`sqrt(x)`	Returns the square root of x.
`to_base(x, radix)`	Returns the radix representation of x.
`truncate(x)`	Returns x rounded to integer by dropping digits after decimal point.
`acos(x)`	Returns the arc cosine of x.
`asin(x)`	Returns the arc sine of x.
`atan(x)`	Returns the arc tangent of x.
`atan2(y,x)`	Returns the arc tangent of y/x.
`cos(x)`	Returns the cosine of x.
`sin(x)`	Returns the sine of x.
`cosh(x)`	Returns the hyperbolic cosine of x.

Function	Description
`tan(x)`	Returns the tangent of x.
`tanh(x)`	Returns the hyperbolic tangent of x.
`infinity()`	Returns the constant representing positive infinity.
`is_infinity(x)`	Determines whether x is infinite.
`is_finity(x)`	Determines whether x is finite.
`is_nan(x)`	Determines whether x is not-a-number.

18.9.6.6 String functions

The query and analysis feature of Log Service supports string functions.

The query and analysis feature of Log Service allows you to use string functions for log analysis. The following table describes the supported string functions.

Function	Description
`length(x)`	Returns the length of x.
`levenshtein_distance(string1, string2)`	Returns the minimum edit distance of string1 and string2.
`lower(string)`	Converts string to lowercase.
`ltrim(string)`	Removes leading whitespaces from string.
`replace(string, search)`	Removes all instances of search from string.
`replace(string, search,rep)`	Replaces all instances of search with rep in string.
`reverse(string)`	Returns string with the characters in reverse order.
`rtrim(string)`	Removes trailing whitespaces from string.
`split(string,delimiter,limit)`	Splits string on delimiter and returns an array of size at most limit. Values of limit start with 1.
`split_part(string,delimiter,offset)`	Splits string on delimiter and returns the array specified with offset. Values of offset start with 1.

Function	Description
`strpos(string, substring)`	Returns the starting position of the first instance of substring in string. Positions start with 1. If not found, 0 is returned.
`substr(string, start)`	Returns the rest of string from the starting position start. Positions start with 1.
`substr(string, start, length)`	Returns a substring from string of the specified length from the starting position start. Positions start with 1.
`trim(string)`	Removes leading and trailing whitespaces from string.
`upper(string)`	Converts string to uppercase.
`concat(string,string.....)`	Concatenates two or more strings into a single string.
`hamming_distance (string1,string2)`	Returns the Hamming distance of string1 and string2.

Note:

Strings must be enclosed in single quotation marks, and double quotation marks indicate column names. For example, a='abc' means column a = string abc, and a = "abc" means column a = column abc.

18.9.6.7 Date and time functions

Log Service supports date and time functions. You can apply the following date and time functions to analysis statements.

Date and time types

- **unixtime: indicates the number of seconds that have elapsed since January 1, 1970. For example, `1512374067` is equivalent to `Mon Dec 4 15:54:27 CST 2017`. In Log Service, the value of the `__time__` field in each log is of this type.**
- **timestamp: indicates the time in string format, such as `2017-11-01 13:30:00`.**

Date functions

Log Service supports the following common date functions.

Function	Description	Example
`current_date`	**Returns the current date.**	`latency>100| select current_date`
`current_time`	**Returns the current time in format of hour:minute :second:millisecond time zone.**	`latency>100| select current_time`
`current_timestamp`	**Returns the current timestamp. It is the combination of current_date and current_time.**	`latency>100| select current_timestamp`
`current_timezone()`	**Returns the current time zone.**	`latency>100| select current_timezone()`
`from_iso8601_timestamp( string)`	**Parses the ISO 8601 formatted string into a timestamp with time zone .**	`latency>100| select from_iso8601_timestamp( iso8601)`
`from_iso8601_date( string)`	**Parses the ISO 8601 formatted string into a date.**	`latency>100| select from_iso8601_date( iso8601)`
`from_unixtime(unixtime)`	**Returns the UNIX timestamp unixtime as a timestamp.**	`latency>100| select from_unixtime( 1494985275)`
`from_unixtime(unixtime, string)`	**Returns the UNIX timestamp unixtime as a timestamp with time zone using string.**	`latency>100| select from_unixtime ( 1494985275,'Asia/ Shanghai')`
`localtime`	**Returns the current time.**	`latency>100| select localtime`
`localtimestamp`	**Returns the current timestamp.**	`latency>100| select localtimestamp`
`now()`	**Returns the current date and time. This is an alias for `current_timestamp`.**	**N/A**

Function	Description	Example
`to_unixtime(timestamp)`	**Returns timestamp as a UNIX timestamp.**	`*| select to_unixtime('2017-05-17 09:45:00.848 Asia/Shanghai')`

Time functions

MySQL time formats

Log Service supports MySQL time formats such as %a, %b, and %y.

Function	Description	Example
`date_format(timestamp, format)`	**Formats timestamp as a string using format.**	`latency>100| select date_format (date_parse ('2017-05-17 09:45:00 ','%Y-%m-%d %H:%i:%S'), '%Y-%m-%d') group by method`
`date_parse(string, format)`	**Parses string into a timestamp using format.**	`latency>100|select date_parse('2017-05-17 09:45:00','%Y-%m-%d %H:%i:%S') group by method`

Table 18-9: Format description

Format	Description
%a	**The abbreviated week day name, such as Sun and Sat.**
%b	**The abbreviated month name, such as Jan and Dec.**
%c	**The month in numeric format. Valid values: 1 to 12.**
%D	**The day of the month with English suffix, such as 0th, 1st, 2nd, and 3rd.**
%d	**The day of the month in numeric format. Valid values: 01 to 31.**
%e	**The day of the month in numeric format. Valid values: 1 to 31.**
%H	**The hour in the 24-hour format.**
%h	**The hour in the 12-hour format.**
%I	**The hour in the 12-hour format.**

Format	Description
%i	The minute in numeric format. Valid values: 00 to 59.
%j	The day of the year. Valid values: 001 to 366.
%k	The hour. Valid values: 0 to 23.
%l	The hour. Valid values: 1 to 12.
%M	The month name. Valid values: January to December.
%m	The month in numeric format. Valid values: 01 to 12.
%p	The a.m. or p.m..
%r	The time in 12-hour format: `hh:mm:ss AM/PM`.
%S	The second. Valid values: 00 to 59.
%s	The second. Valid values: 00 to 59.
%T	The time in 24-hour format: `hh:mm:ss`.
%U	The week of the year. Sunday is the first day of the week. Valid values: 00 to 53.
%u	The week of the year. Monday is the first day of the week. Valid values: 00 to 53.
%V	The week of the year. Sunday is the first day of the week. This format is used together with %X. Valid values: 01 to 53.
%v	The week of the year. Monday is the first day of the week. This format is used together with %x. Valid values: 01 to 53.
%W	The week day name. Valid values: Sunday to Saturday.
%w	The day of the week. The valid value ranges from 0 to 6 . The value 0 indicates Sunday.
%Y	The year in 4-digit format.
%y	The year in 2-digit format.
%%	The % escape character.

Truncation functions

Log Service supports a truncation function, which can return a time truncated by second, minute, hour, day, month, or year. Typically, this function is used in scenarios where time-based statistics are performed.

Function syntax:

```
date_trunc(unit, x)
```

Arguments:

The following table lists the values for unit (x is `2001-08-22 03:04:05.000`).

Unit	Converted result
second	2001-08-22 03:04:05.000
minute	2001-08-22 03:04:00.000
hour	2001-08-22 03:00:00.000
day	2001-08-22 00:00:00.000
week	2001-08-20 00:00:00.000
month	2001-08-01 00:00:00.000
quarter	2001-07-01 00:00:00.000
year	2001-01-01 00:00:00.000

x can be of the timestamp or unixtime type.

date_trunc is only applicable to scenarios where statistics are collected at a fixed interval. For scenarios where statistics are collected based on flexible time dimensions, for example, every 5 minutes, perform GROUP BY based on the modulo method.

```
* | SELECT count(1) as pv,  __time__ - __time__% 300 as minute5groupby
 minute5 limit 100
```

In the preceding formula, `%300` indicates that the modulo operation is performed every 5 minutes when collecting statistics.

Date function example

The following comprehensive example is a date function that uses the time format:

```
*|select  date_trunc('minute' ,  __time__)  as t,
       truncate (avg(latency) ) ,
       current_date
       group by   t
       order by t  desc
```

```
        limit 60
```

18.9.6.8 URL functions

The query and analysis feature of Log Service supports URL functions.

URL functions extract fields from standard URLs. A standard URL is as follows:

```
[protocol:][//host[:port]][path][? query][#fragment]
```

Common URL functions

Function	Description	Example
`url_extract_fragment(url)`	**Returns the fragment identifier from url. The result is of the varchar type.**	`*| select url_extract_fragment(url)`
`url_extract_host(url)`	**Returns the host from url. The result is of the varchar type.**	`*|select url_extract_host(url)`
`url_extract_parameter(url, name)`	**Returns the value of the first query string parameter named name from url. The result is of the varchar type.**	`*|select url_extract_parameter(url)`
`url_extract_path(url)`	**Returns the path from url. The result is of the varchar type.**	`*|select url_extract_path(url)`
`url_extract_port(url)`	**Returns the port number from url. The result is of the bigint type.**	`*|select url_extract_port(url)`
`url_extract_protocol(url)`	**Returns the protocol from url. The result is of the varchar type.**	`*|select url_extract_protocol(url)`
`url_extract_query(url)`	**Returns the query string from url. The result is of the varchar type.**	`*|select url_extract_query(url)`
`url_encode(value)`	**Escapes the URL value by encoding it.**	`*|select url_encode(url)`
`url_decode(value)`	**Decodes the URL value.**	`*|select url_decode(url)`

18.9.6.9 Regular expression functions

The query and analysis feature of Log Service supports regular expression functions.

A regular expression function parses a string and returns the required substrings.

The following table describes common regular expression functions.

Function	Description	Example
`regexp_extract_all( string, pattern)`	**Returns substrings matched by the regular expression pattern in string.**	`|*SELECT regexp_extract_all('5a 67b 890m', '\d+')`: **returns** `['5','67','890']`. `*|SELECT regexp_extract_all('5a 67a 890m', '(\d+)a')`: **returns** `['5a','67a']`.
`regexp_extract_all( string, pattern, group)`	**Finds all occurrences of the regular expression pattern in string and returns the capturing group number group.**	`*| ` SELECT regexp_extract_all(‘5a 67a 890m’, ‘(\d+)a’,1)`: **returns** `[‘5’,’67’]`.
`regexp_extract(string, pattern)`	**Returns the first substring matched by the regular expression pattern in string.**	`*|SELECT regexp_extract('5a 67b 890m', '\d+')`: **returns** `'5'`.
`regexp_extract(string, pattern,group)`	**Finds the first occurrence of the regular expression pattern in string and returns the capturing group number group.**	`*|SELECT regexp_extract('5a 67b 890m', '(\d+)([a-z]+)',2)`: **returns** `'b'`.
`regexp_like(string, pattern)`	**Evaluates the regular expression pattern and determines whether it is contained within string. A Boolean value is returned.**	`*|SELECT regexp_like('5a 67b 890m', '\d+m')`: **returns true.**
`regexp_replace(string, pattern, replacement)`	**Replaces every instance of the substring matched by the regular expression pattern in string with replacement.**	`*|SELECT regexp_replace('5a 67b 890m', '\d+','a')`: **returns** `'aa ab am'`.

Function	Description	Example
`regexp_replace(string, pattern)`	**Removes every instance of the substring matched by the regular expression pattern from string. This function is equivalent to `regexp_replace(string, pattern,'')`.**	`*|SELECT regexp_replace ('5a 67b 890m', '\d+')`: **returns `'a b m'`.**
`regexp_split(string, pattern)`	**Splits string by using the regular expression pattern and returns an array.**	`*|SELECT regexp_split ('5a 67b 890m', '\d+')`: **returns `['a','b','m']`.**

18.9.6.10 JSON functions

The query and analysis feature of Log Service supports JSON functions.

JSON functions can parse a string as the JSON type and extract the fields in JSON. JSON mainly has the following two structures: map and array. If a string fails to be parsed as the JSON type, null is returned.

Log Service supports the following JSON functions.

Function	Description	Example
`json_parse(string)`	**Returns the JSON value deserialized from the input JSON string.**	`SELECT json_parse('[1, 2, 3]')` **: returns a JSON array**
`json_format(json)`	**Returns the JSON string serialized from the input JSON value.**	`SELECT json_format( json_parse('[1, 2, 3 ]'))` **: returns a string**
`json_array_contains( json, value)`	**Determines whether a value exists in json (a string containing a JSON array).**	`SELECT json_array _contains(json_parse('[ 1, 2, 3]'), 2)` **or** `SELECT json_array_contains('[ 1, 2, 3]', 2)`
`json_array_get( json_array, index)`	**Returns the element at the specified index into the json_array. This function is equivalent to `json_array_contains`.**	`SELECT json_array_get ('["a", "b", "c"]', 0)`: **returns 'a'**

Function	Description	Example
`json_array_length(json)`	**Returns the array length of json (a string containing a JSON array):**	`SELECT json_array_length('[1, 2, 3]')`: **returns 3**
`json_extract(json, json_path)`	**Evaluates the JSONPath-like expression json_path on json (a string containing JSON) and returns the result as a JSON string. The json_path syntax is similar to** `$.store.book[0].title`**.**	`SELECT json_extract(json, '$.store.book');`
`json_extract_scalar(json, json_path)`	**Returns the result as a string (as opposed to being encoded as JSON by** `json_extract`**).**	**N/A**
`json_size(json, json_path)`	**Returns the size of a JSON object or array.**	`SELECT json_size('[1, 2, 3]')`: **returns 3**

18.9.6.11 Type conversion functions

The query and analysis feature of Log Service supports type conversion functions.

Log Service supports data types such as long, double, and text in configurations and data types such as bigint, double, varchar, timestamp, and int in queries.

The following type conversion function explicitly converts values in a column to a specified type:

```
try_cast(value AS type) → type
```

18.9.6.12 GROUP BY syntax

The query and analysis feature of Log Service supports the GROUP BY syntax.

GROUP BY supports multiple columns and can use a SELECT column alias to indicate the corresponding KEY.

Example:

```
method:PostLogstoreLogs |select avg(latency),projectName,date_trunc('
hour',__time__) as hour   group by projectName,hour
```

The alias hour indicates the third SELECT column `date_trunc('hour',__time__)`.

This is very helpful for complex queries.

GROUP BY supports GROUPING SETS, CUBE, and ROLLUP.

Example:

```
method:PostLogstoreLogs |select avg(latency)   group by cube(
projectName,logstore)
method:PostLogstoreLogs |select avg(latency)   group by GROUPING SETS
 ( ( projectName,logstore), (projectName,method))
method:PostLogstoreLogs |select avg(latency)   group by rollup(
projectName,logstore)
```

Examples

Perform GROUP BY according to time

Each log has a built-in time column `__time__`. When the statistical function of any column is enabled, the statistics of the time column will be included.

The `date_trunc` function can truncate the time column to minute, hour, day, month, and year. `date_trunc` accepts a truncation unit and a column of UNIX time or timestamp type, such as `__time__`.

- **PV statistics per hour and per minute**

```
* | SELECT count(1) as pv , date_trunc('hour',__time__) as hour
group by hour order by hour limit 100
 * | SELECT count(1) as pv , date_trunc('minute',__time__) as minute
 group by minute order by minute limit 100
```

Note:

limit 100 indicates that up to 100 rows are retrieved. If the LIMIT clause is not added, up to 10 rows of data are retrieved by default.

- **date_trunc is only applicable to statistics at a fixed time interval. For statistics based on flexible time dimensions, for example, every 5 minutes, perform GROUP BY in mod.**

  ```
  * | SELECT count(1) as pv,  __time__ - __time__% 300 as minute5
  group by minute5 limit 100
  ```

  In the preceding formula, `%300` indicates that the time is truncated in mod every 5 minutes.

Retrieve non-aggregation columns in GROUP BY

In standard SQL, if the GROUP BY syntax is used during the SELECT operation, the system only selects the raw data of the SELECT GROUP BY column, or performs aggregation on any columns. Retrieving data from non-GROUP BY columns is not allowed.

For example, the following syntax is invalid. Because b is a non-GROUP BY column, the system cannot determine which row of b to return during GROUP BY based on a.

```
*|select a, b , count(c) group by a
```

Instead, you can use the `arbitrary` function to return b.

```
*|select a, arbitrary(b), count(c) group by a
```

18.9.6.13 Window functions

The query and analysis feature of Log Service supports window functions.

Window functions perform calculations across rows of a query result. Common SQL aggregate functions calculate the results of only one row or aggregate all rows into one row for calculation. Window functions support cross-row calculation and fill the calculation results in each row.

Syntax of window functions:

```
SELECT key1, key2, value,
       rank() OVER (PARTITION BY key2
                    ORDER BY value DESC) AS rnk
FROM orders
```

```
ORDER BY key1,rnk
```

Key part:

```
 rank() OVER (PARTITION BY  KEY1 ORDER BY KEY2 DESC)
```

rank() is an aggregate function. You can use any function in analysis syntax or the function listed in this topic. PARTITION BY indicates the buckets based on which values are calculated.

Special aggregate functions used in windows

Function	Description
rank()	Returns the rank of a value in a group of values. The rank is one plus the number of rows preceding the row that are not peer with the row.
row_number()	Returns a unique, sequential number for each row.
first_value(x)	Returns the first value of the window. Typically, the function is used to obtain the maximum value after values of the window are sorted.
last_value(x)	Returns the last value of the window.
nth_value(x, offset)	Returns the value at the specified offset from the beginning of the window.
lead(x,offset,defaut_value)	Returns the value at offset rows after the current row in the window. If the target row does not exist, the default_value is returned.
lag(x,offset,defaut_value)	Returns the value at offset rows before the current row in the window. If the target row does not exist, the default_value is returned.

Example

- **Rank the salaries of employees in their respective departments**

```
 * | select department, persionId, salary , rank() over(PARTITION
 BY department order by salary desc) as salary_rank  order by
department,salary_rank
```

Response results

department	persionId	salary	salary_rank
dev	john	9000	1
dev	Smith	8000	2
dev	Snow	7000	3
dev	Achilles	6000	4
Marketing	Blan Stark	9000	1
Marketing	Rob Stark	8000	2
Marketing	Sansa Stark	7000	3

- **Calculate the salaries of employees as percentages in their respective departments**

```
* | select department, persionId, salary *1.0 / sum(salary) over(
PARTITION BY department ) as salary_percentage
```

Response results

department	persionId	salary	salary_percentage
dev	john	9000	0.3
dev	Smith	8000	0.26
dev	Snow	7000	0.23
dev	Achilles	6000	0.2
Marketing	Blan Stark	9000	0.375
Marketing	Rob Stark	8000	0.333
Marketing	Sansa Stark	7000	0.29

- **Calculate the daily UV increase over the previous day**

```
* | select day ,uv, uv *1.0 /(lag(uv,1,0) over() ) as diff_perce
ntage from
(
select approx_distinct(ip) as uv, date_trunc('day',__time__) as day
from log group by day order by day asc
```

```
)
```

Response results

day	uv	diff_percentage
2017-12-01 00:00:00	100	null
2017-12-02 00:00:00	125	1.25
2017-12-03 00:00:00	150	1.2
2017-12-04 00:00:00	175	1.16
2017-12-05 00:00:00	200	1.14
2017-12-06 00:00:00	225	1.125
2017-12-07 00:00:00	250	1.11

18.9.6.14 HAVING syntax

The query and analysis feature of Log Service supports the HAVING syntax of standard SQL. The HAVING syntax is used together with the GROUP BY syntax to filter GROUP BY results.

Format:

```
method :PostLogstoreLogs |select avg(latency),projectName group by
projectName  HAVING avg(latency) > 100
```

Difference between HAVING and WHERE

HAVING filters the aggregated results grouped by using GROUP BY, and WHERE filters raw data during data aggregation.

Example

Calculate the average rainfall of each province where temperature is above 10°C , and only display the provinces with average rainfall above 100 mL in the final results:

```
* | select avg(rain) ,province where teporature > 10groupby province
having avg(rain) > 100
```

18.9.6.15 ORDER BY syntax

The query and analysis feature of Log Service supports the ORDER BY syntax.

ORDER BY is used to sort the output results. Currently, you can only sort the results by one column.

Syntax format:

```
order by column name [desc|asc]
```

Example:

```
method :PostLogstoreLogs |select avg(latency) as avg_latency,
projectName group by projectName
HAVING avg(latency) > 5700000
order by avg_latency desc
```

18.9.6.16 LIMIT syntax

The query and analysis feature of Log Service supports the LIMIT syntax.

LIMIT is followed by a number to restrict the maximum number of rows in output results. If no LIMIT clause is added, only 10 rows are returned by default.

 **Note:**

The LIMIT OFFSET syntax and LINES syntax are not supported.

Example:

```
*| select avg(latency) as avg_latency , methodgroupbymethodorderbyavg_
latencydesclimit100
```

18.9.6.17 CASE WHEN syntax

The query and analysis feature of Log Service supports the CASE WHEN syntax.

The CASE WHEN syntax is used to classify continuous data. For example, you can use the CASE WHEN syntax to extract information from http_user_agent and classify the information into two types: Android and iOS.

```
SELECT
 CASE
```

```
 WHEN http_user_agent like '%android%' then 'android'
 WHEN http_user_agent like '%ios%' then 'ios'
 ELSE 'unknown' END
 as http_user_agent,
    count(1) as pv
    group by  http_user_agent
```

Examples

- **Calculate the proportion of requests with status code 200 to all requests:**

```
* | SELECT
 sum(
 CASE
 WHEN status =200 then 1
 ELSE 0 end
 ) *1.0 / count(1) as status_200_percentage
```

- **Calculate the distribution of latencies:**

```
* | SELECT `
 CASE
 WHEN latency < 10 then 's10'
 WHEN latency < 100 then 's100'
 WHEN latency < 1000 then 's1000'
 WHEN latency < 10000 then 's10000'
 else 's_large' end
 as latency_slot,
 count(1) as pv
 group by latency_slot
```

IF syntax

The IF syntax is logically equivalent to the CASE WHEN syntax.

```
CASE
     WHEN condition THEN true_value
     [ ELSE false_value ]
 END
```

- **if(condition, true_value)**
  - **If the condition is true, true_value is returned. Otherwise, null is returned.**
- **if(condition, true_value, false_value)**
  - **If the condition is true, true_value is returned. Otherwise, the false_value column is returned.**

COALESCE syntax

The COALESCE syntax returns the first non-NULL value from multiple columns.

```
COALESCE (value1, value2 [,...])
```

NULLIF syntax

If value1 equals value2, null is returned. Otherwise, value1 is returned.

```
nullif(value1, value2)
```

TRY syntax

The TRY syntax catches some of the underlying exceptions, such as division by zero errors, and returns null.

```
try(expression)
```

18.9.6.18 Nested queries

The query and analysis function of Log Service supports nested queries.

You can use nested queries to perform more complicated queries.

Nested queries differ from non-nested queries in that you need to specify the FROM clause in the SQL statement. You need to specify the `from log` keyword in the query to read raw data from logs.

Example:

```
* | select sum(pv) from
(
select count(1) as pv from log group by method
)
```

18.9.6.19 Array functions

The query and analysis feature of Log Service supports array functions.

Function	Description	Example
Subscript operator ([])	**Obtains a certain element in an array.**	**N/A**

Function	Description	Example

Concatenation operator (||)	Connects two arrays into one.	`SELECT ARRAY [1] || ARRAY [2]; – [1, 2]`
		`SELECT ARRAY [1] || 2; – [1, 2]`
		`SELECT 2 || ARRAY [1]; – [2, 1]`

array_distinct	Removes duplicate values from an array.	N/A
array_intersect(x, y)	Obtains the intersection of arrays x and y.	N/A
array_union(x, y) → array	Obtains the union of arrays x and y.	N/A
array_except(x, y) → array	Obtains the subtraction of arrays x and y.	N/A
array_join(x, delimiter , null_replacement) → varchar	Concatenates the elements of the given array using the delimiter and an optional string to replace null elements.	N/A
array_max(x) → x	Returns the maximum value of the input array.	N/A
array_min(x) → x	Returns the minimum value of the input array.	N/A
array_position(x, element) → bigint	Returns the position of the first occurrence of the given element in array x (or 0 if not found).	N/A
array_remove(x, element) → array	Removes all elements that equal the given element from array x.	N/A
array_sort(x) → array	Sorts and returns array x. The elements of x must be orderable. Null elements will be placed at the end of the returned array.	N/A

Function	Description	Example
cardinality(x) → bigint	Returns the cardinality (size) of array x.	N/A
concat(array1, array2, ⋯, arrayN) → array	Concatenates the arrays array1, array2, ⋯, and arrayN.	N/A
contains(x, element) → boolean	Returns true if array x contains element.	N/A
filter(array, function) → array	For more information about this function (a Lambda function), see filter in *Lambda functions*.	N/A
flatten(x) → array	Flattens an array(array (T)) to an array(T) by concatenating the contained arrays.	N/A
reduce(array, initialState, inputFunction, outputFunction) → x	For more information about this function, see reduce in *Lambda functions*.	N/A
reverse(x) → array	Returns an array which has the reversed order of array x.	N/A
sequence(start, stop) → array	Generates a sequence of items from start to stop, increasing by 1.	N/A
sequence(start, stop, step) → array	Generates a sequence of items from start to stop, increasing by step.	N/A
sequence(start, stop, step) → array	Generates a sequence of timestamps from start to stop, increasing by step . The type of step can be either INTERVAL DAY TO SECOND or INTERVAL YEAR TO MONTH.	N/A
shuffle(x) → array	Generates a random permutation of array x.	N/A

Function	Description	Example
slice(x, start, length) → array	Returns a subset of array x starting from the start value with the given length.	N/A
transform(array, function) → array	For more information about this function, see transform() in *Lambda functions*.	N/A
zip(array1, array2[, ⋯]) → array	Merges the given arrays. The M-th element of the N-th argument will be the N-th field of the M-th output element.	`SELECT zip(ARRAY[1, 2], ARRAY['1b', null, '3b']); — [ROW(1, '1b'), ROW(2, null), ROW(null, '3b')]`
zip_with(array1, array2, function) → array	For more information about this function, see zip_with in *Lambda functions*.	N/A

18.9.6.20 Binary string functions

The query and analysis feature of Log Service supports binary string functions.

The binary string type varbinary is different from the string type varchar.

Function	Description
Concatenation operator (||)	Concatenates strings. The result of `a || b` is `ab`.
length(binary) → bigint	Returns the length of a binary string in bytes.
concat(binary1, ⋯, binaryN) → varbinary	Returns the concatenation of binary1, ..., binaryN. This function provides the same functionality as the SQL-standard concatenation operator (||).
to_base64(binary) → varchar	Encodes a binary string into a Base64 encoded string representation.
from_base64(string) → varbinary	Decodes a binary string from a Base64 encoded string.
to_base64url(binary) → varchar	Encodes a binary string into a Base64 encoded string representation using the URL safe alphabet.

Function	Description
from_base64url(string) → varbinary	Decodes a binary string from a Base64 encoded string using the URL safe alphabet.
to_hex(binary) → varchar	Encodes a binary string into a hexadecimal string representation.
from_hex(string) → varbinary	Decodes a binary string from a hexadecimal encoded string.
to_big_endian_64(bigint) → varbinary	Encodes a bigint value in 64-bit two's complement big endian format.
from_big_endian_64(binary) → bigint	Decodes a bigint value from a 64-bit two's complement big endian binary string.
md5(binary) → varbinary	Computes the MD5 hash of a binary string.
sha1(binary) → varbinary	Computes the SHA1 hash of a binary string.
sha256(binary) → varbinary	Computes the SHA256 hash of a binary string.
sha512(binary) → varbinary	Computes the SHA512 hash of a binary string.
xxhash64(binary) → varbinary	Computes the xxhash64 hash of a binary string.

18.9.6.21 Bitwise functions

The query and analysis feature of Log Service allows you to use bitwise functions for log analysis.

Function	Description	Example

bit_count(x, bits) → bigint	**Counts the number of bits set in x (treated as a signed integer with the given number of bits) in two's complement.**	`SELECT bit_count(9, 64); – 2`
		`SELECT bit_count(9, 8); – 2`
		`SELECT bit_count(-7, 64); – 62`
		`SELECT bit_count(-7, 8); – 6`

bitwise_and(x, y) → bigint	**Returns the bitwise AND of x and y in two's complement.**	**N/A**
bitwise_not(x) → bigint	**Returns the bitwise NOT of x in two's complement.**	**N/A**
bitwise_or(x, y) → bigint	**Returns the bitwise OR of x and y in two's complement.**	**N/A**
bitwise_xor(x, y) → bigint	**Returns the bitwise XOR of x and y in two's complement.**	**N/A**

18.9.6.22 Comparison functions and operators

The query and analysis feature of Log Service supports comparison functions and operators.

Comparison functions and operators

A comparison function compares two values, and is applicable to any comparable types of data, such as int, bigint, double, and text data.

Comparison operators

A comparison operator is used to compare two values. If the statement is true, TRUE is returned. Otherwise, FALSE is returned.

Operator	Description
<	Less than
>	Greater than
<=	Less than or equal to
> =	Greater than or equal to
=	Equal to
<>	Not equal to
! =	Not equal to

Range operators

The BETWEEN operator is used to determine whether a value is within a specified range.

- If the value is within the specified range, TRUE is returned. Otherwise, FALSE is returned.

  Example: `SELECT 3 BETWEEN 2 AND 6`. The statement is true, and TRUE is returned.

  The preceding statement is equivalent to `SELECT 3 >= 2 AND 3 <= 6`.
- The BETWEEN operator can follow the NOT operator to test whether a value is not within a specified range.

  Example: `SELECT 3 NOT BETWEEN 2 AND 6`. The statement is false, and FALSE is returned.

  The preceding statement is equivalent to `SELECT 3 < 2 OR 3 > 6`.
- If any of the three values is NULL, NULL is returned.

IS NULL and IS NOT NULL

The IS NULL and IS NOT NULL operators test whether a value is NULL.

IS DISTINCT FROM and IS NOT DISTINCT FROM

These operators are similar to the comparison operators EQUAL TO and NOT EQUAL TO, but they can determine whether a NULL value exists.

Examples:

```
SELECT NULL IS DISTINCT FROM NULL -- false
```

```
SELECT NULL IS NOT DISTINCT FROM NULL -- true
```

DISTINCT can be used to compare parameter values under multiple conditions, as described in the following table.

a	b	a = b	a <> b	a DISTINCT b	a NOT DISTINCT b
1	1	TRUE	FALSE	FALSE	TRUE
1	2	FALSE	TRUE	TRUE	FALSE
1	NULL	NULL	NULL	TRUE	FALSE
NULL	NULL	NULL	NULL	FALSE	TRUE

GREATEST and LEAST

These operators are used to obtain the maximum or minimum values across multiple columns.

Examples:

```
select greatest(1,2,3) -- Returns 3.
```

Quantified comparison predicates: ALL, ANY, and SOME

The ALL, ANY, and SOME quantifiers can be used to determine whether a parameter meets the specified conditions.

- **ALL is used to determine whether a parameter meets all the conditions. If the statement is true, TRUE is returned. Otherwise, FALSE is returned.**
- **ANY is used to determine whether a parameter meets any of the conditions. If the statement is true, TRUE is returned. Otherwise, FALSE is returned.**
- **Same as ANY, SOME is used to determine whether a parameter meets any of the conditions.**
- **ALL, ANY, and SOME must immediately follow comparison operators.**

ALL and ANY support comparison under multiple conditions, as described in the following table.

Expression	Description
A = ALL (…)	Evaluates to TRUE when A is equal to all values.

Expression	Description
A <> ALL (…)	Evaluates to TRUE when A does not match any value.
A < ALL (…)	Evaluates to TRUE when A is smaller than the smallest value.
A = ANY (…)	Evaluates to TRUE when A is equal to any of the values. This form is equivalent to A IN (...).
A <> ANY (…)	Evaluates to TRUE when A does not match one or more values.
A < ANY (…)	Evaluates to TRUE when A is smaller than the greatest value.

Examples:

```
SELECT 'hello' = ANY (VALUES 'hello', 'world'); -- true
SELECT 21 < ALL (VALUES 19, 20, 21); -- false
SELECT 42 >= SOME (SELECT 41 UNION ALL SELECT 42 UNION ALL SELECT 43);
 -- true
```

18.9.6.23 Lambda functions

The query and analysis feature of Log Service supports Lambda functions.

Lambda expressions

Lambda expressions are written with

```
->
```

Examples:

```
x -> x + 1
(x, y) -> x + y
x -> regexp_like(x, 'a+')
x -> x[1] / x[2]
x -> IF(x > 0, x, -x)
x -> COALESCE(x, 0)
x -> CAST(x AS JSON)
x -> x + TRY(1 / 0)
```

Most MySQL expressions can be used in a Lambda body.

filter(array<T>, function<T, boolean>) → ARRAY<T>

Constructs an array from those elements of the given array for which function returns true.

Examples:

```
SELECT filter(ARRAY [], x -> true); -- []
SELECT filter(ARRAY [5, -6, NULL, 7], x -> x > 0); -- [5, 7]
SELECT filter(ARRAY [5, NULL, 7, NULL], x -> x IS NOT NULL); -- [5, 7]
```

map_filter(map<K, V>, function<K, V, boolean>) → MAP<K,V>

Constructs a map from those entries of the given map for which function returns true.

Examples:

```
SELECT map_filter(MAP(ARRAY[], ARRAY[]), (k, v) -> true); -- {}
SELECT map_filter(MAP(ARRAY[10, 20, 30], ARRAY['a', NULL, 'c']), (k, v
) -> v IS NOT NULL); -- {10 -> a, 30 -> c}
SELECT map_filter(MAP(ARRAY['k1', 'k2', 'k3'], ARRAY[20, 3, 15]), (k,
v) -> v > 10); -- {k1 -> 20, k3 -> 15}
```

reduce(array<T>, initialState S, inputFunction<S, T, S>, outputFunction<S, R>) → R

The reduce function retrieves each element in the array from the initial state, calculates inputFunction(s,t) based on the state S, and generates a new state. The final step invokes outputFunction to turn the final state into the result R.

1. **Starts from the initial state S.**
2. **Retrieves each element T.**
3. **Calculates inputFunction(S,T) to generate a new state S.**
4. **Repeats steps 2 and 3 to the last element.**
5. **Turns the final state into the result R.**

Examples:

```
SELECT reduce(ARRAY [], 0, (s, x) -> s + x, s -> s); -- 0
SELECT reduce(ARRAY [5, 20, 50], 0, (s, x) -> s + x, s -> s); -- 75
SELECT reduce(ARRAY [5, 20, NULL, 50], 0, (s, x) -> s + x, s -> s);
 -- NULL
SELECT reduce(ARRAY [5, 20, NULL, 50], 0, (s, x) -> s + COALESCE(x, 0
), s -> s); -- 75
SELECT reduce(ARRAY [5, 20, NULL, 50], 0, (s, x) -> IF(x IS NULL, s, s
 + x), s -> s); -- 75
SELECT reduce(ARRAY [2147483647, 1], CAST (0 AS BIGINT), (s, x) -> s
 + x, s -> s); -- 2147483648
SELECT reduce(ARRAY [5, 6, 10, 20], -- calculates arithmetic average:
10.25
              CAST(ROW(0.0, 0) AS ROW(sum DOUBLE, count INTEGER)),
              (s, x) -> CAST(ROW(x + s.sum, s.count + 1) AS ROW(sum
DOUBLE, count INTEGER)),
```

```
                s -> IF(s.count = 0, NULL, s.sum / s.count));
```

transform(array<T>, function<T, U>) → ARRAY<U>

Returns an array that is the result of applying function to each element of the given array.

Examples:

```
SELECT transform(ARRAY [], x -> x + 1); -- []
SELECT transform(ARRAY [5, 6], x -> x + 1); -- [6, 7] -- Increments
each element by 1.
SELECT transform(ARRAY [5, NULL, 6], x -> COALESCE(x, 0) + 1); -- [6,
1, 7]
SELECT transform(ARRAY ['x', 'abc', 'z'], x -> x || '0'); -- ['x0', '
abc0', 'z0']
SELECT transform(ARRAY [ARRAY [1, NULL, 2], ARRAY[3, NULL]], a ->
filter(a, x -> x IS NOT NULL)); -- [[1, 2], [3]]
```

transform_keys(map<K1, V>, function<K1, V, K2>) → MAP<K2,V>

Returns a map that applies function to each entry of the given map and transforms the keys.

Examples:

```
SELECT transform_keys(MAP(ARRAY[], ARRAY[]), (k, v) -> k + 1); -- {}
SELECT transform_keys(MAP(ARRAY [1, 2, 3], ARRAY ['a', 'b', 'c']), (k
, v) -> k + 1); -- {2 -> a, 3 -> b, 4 -> c}  -- Increments each key by
 1.
SELECT transform_keys(MAP(ARRAY ['a', 'b', 'c'], ARRAY [1, 2, 3]), (k
, v) -> v * v); -- {1 -> 1, 4 -> 2, 9 -> 3}
SELECT transform_keys(MAP(ARRAY ['a', 'b'], ARRAY [1, 2]), (k, v) -> k
 || CAST(v as VARCHAR)); -- {a1 -> 1, b2 -> 2}
SELECT transform_keys(MAP(ARRAY [1, 2], ARRAY [1.0, 1.4]), -- {one ->
1.0, two -> 1.4}
                      (k, v) -> MAP(ARRAY[1, 2], ARRAY['one', 'two'])[
k]);
```

transform_values(map<K, V1>, function<K, V1, V2>) → MAP<K, V2>

Returns a new map `<K,V2>` that applies function to each entry of the given map and transforms the values.

```
SELECT transform_values(MAP(ARRAY[], ARRAY[]), (k, v) -> v + 1); -- {}
SELECT transform_values(MAP(ARRAY [1, 2, 3], ARRAY [10, 20, 30]), (k,
v) -> v + 1); -- {1 -> 11, 2 -> 22, 3 -> 33}
SELECT transform_values(MAP(ARRAY [1, 2, 3], ARRAY ['a', 'b', 'c']), (
k, v) -> k * k); -- {1 -> 1, 2 -> 4, 3 -> 9}
SELECT transform_values(MAP(ARRAY ['a', 'b'], ARRAY [1, 2]), (k, v) -
> k || CAST(v as VARCHAR)); -- {a -> a1, b -> b2}
SELECT transform_values(MAP(ARRAY [1, 2], ARRAY [1.0, 1.4]), -- {1 ->
one_1.0, 2 -> two_1.4}
```

```
                          (k, v) -> MAP(ARRAY[1, 2], ARRAY['one', 'two
'])[k] || '_' || CAST(v AS VARCHAR));
```

zip_with(array<T>, array<U>, function<T, U, R>) → array<R>

Merges the two given arrays, element-wise, into a single array using function. Element T in the first array and element U in the second array are used to generate the new array R.

Examples:

```
SELECT zip_with(ARRAY[1, 3, 5], ARRAY['a', 'b', 'c'], (x, y) -> (y, x
)) --Transposes the elements of the two arrays to generate a new array
. Result: [ROW('a', 1), ROW('b', 3), ROW('c', 5)]
SELECT zip_with(ARRAY[1, 2], ARRAY[3, 4], (x, y) -> x + y) -- Result:
 [4, 6]
SELECT zip_with(ARRAY['a', 'b', 'c'], ARRAY['d', 'e', 'f'], (x, y)
 -> concat(x, y)) -- Concatenates the elements of the two arrays to
generate a new string. Result: ['ad', 'be', 'cf']
```

map_zip_with(map<K, V1>, map<K, V2>, function<K, V1, V2, V3>) → map<K, V3>

Merges the two given maps into a single map by applying function to the pair of values with the same key. Values V1 and V2 are used to generate V3 and a new map `<K,V3>` is generated.

```
SELECT map_zip_with(MAP(ARRAY[1, 2, 3], ARRAY['a', 'b', 'c']),
                     MAP(ARRAY[1, 2, 3], ARRAY['d', 'e', 'f']),
                    (k, v1, v2) -> concat(v1, v2)). -- Merges values
with the same key. -- {1 -> ad, 2 -> be, 3 -> cf}
SELECT map_zip_with(MAP(ARRAY['k1', 'k2'], ARRAY[1, 2]),
                    MAP(ARRAY['k2', 'k3'], ARRAY[4, 9]),
                    (k, v1, v2) -> (v1, v2)) -- Generates an array by
 using values v1 and v2. -- {k1 -> ROW(1, null), k2 -> ROW(2, 4), k3 -
> ROW(null, 9)}
SELECT map_zip_with(MAP(ARRAY['a', 'b', 'c'], ARRAY[1, 8, 27]),
                    MAP(ARRAY['a', 'b', 'c'], ARRAY[1, 2, 3]),
                    (k, v1, v2) -> k || CAST(v1/v2 AS VARCHAR)) --
Concatenates the key values and division results of the two values --
 {a -> a1, b -> b4, c -> c9}
```

18.9.6.24 Logical functions

The query and analysis feature of Log Service supports logical functions.

Logical operators

Table 18-10: Logical operators

Operator	Description	Example
AND	The result is TRUE if both values are TRUE.	a AND b

Operator	Description	Example
OR	The result is TRUE if either value is TRUE.	a OR b
NOT	The result is TRUE if the value is FALSE.	NOT a

Effect of NULL on logical operators

The following tables list the truth values when the values of a and b are TRUE, FALSE, are NULL, respectively.

Table 18-11: Truth table 1

a	b	a AND b	a OR b
TRUE	TRUE	TRUE	TRUE
TRUE	FALSE	FALSE	TRUE
TRUE	NULL	NULL	TRUE
FALSE	TRUE	FALSE	TRUE
FALSE	FALSE	FALSE	FALSE
FALSE	NULL	FALSE	NULL
NULL	TRUE	NULL	TRUE
NULL	FALSE	FALSE	NULL
NULL	NULL	NULL	NULL

Table 18-12: Truth table 2

a	NOT a
TRUE	FALSE
FALSE	TRUE
NULL	NULL

18.9.6.25 Column aliases

The query and analysis feature of Log Service allows you to set column aliases.

Context

The SQL standard specifies that a column name must start with a letter and can contain letters, digits, and underscores (_).

If you have configured a column name that does not conform to the SQL standard (such as User-Agent), you need to specify an alias for the column on the statistic properties configuration page. The alias is only used for SQL statistics. The original name is used in underlying storage. Therefore, you must use the original name when performing a search.

In addition, you can give the column an alias to replace the original name for query when the column name is long.

Table 18-13: Sample aliases

Original column name	Alias
User-Agent	ua
User.Agent	ua
123	col
abceefghijklmnopqrstuvw	a

18.9.6.26 Geospatial functions

The query and analysis feature of Log Service supports geospatial functions.

Concept of geometry

Geospatial functions support geometries in the Well-Known Text (WKT) format.

Table 18-14: Geometry formats

Geometry	WKT format
Point	`POINT (0 0)`
LineString	`LINESTRING (0 0, 1 1, 1 2)`
Polygon	`POLYGON ((0 0, 4 0, 4 4, 0 4, 0 0), (1 1, 2 1, 2 2, 1 2, 1 1))`
MultiPoint	`MULTIPOINT (0 0, 1 2)`
MultiLineString	`MULTILINESTRING ((0 0, 1 1, 1 2), (2 3, 3 2, 5 4))`
MultiPolygon	`MULTIPOLYGON (((0 0, 4 0, 4 4, 0 4, 0 0), (1 1, 2 1, 2 2, 1 2, 1 1)), ((-1 -1, -1 -2, -2 -2, -2 -1, -1 -1)))`

Geometry	WKT format
GeometryCollection	`GEOMETRYCOLLECTION (POINT(2 3), LINESTRING (2 3, 3 4))`

Constructors

Table 18-15: Constructor description

Function	Description
ST_Point(double, double) → Point	Returns a geometry type point object with the given coordinate values.
ST_LineFromText(varchar) → LineString	Returns a geometry type linestring object from a WKT representation.
ST_Polygon(varchar) → Polygon	Returns a geometry type polygon object from a WKT representation.
ST_GeometryFromText(varchar) → Geometry	Returns a geometry type object from a WKT representation.
ST_AsText(Geometry) → varchar	Returns the WKT representation of a geometry.

Operations

Function	Description
ST_Boundary(Geometry) → Geometry	Returns the closure of the combinatorial boundary of a geometry.
ST_Buffer(Geometry, distance) → Geometry	Returns the geometry that represents all points whose distance from the specified geometry is less than or equal to the specified distance.
ST_Difference(Geometry, Geometry) → Geometry	Returns the geometry value that represents the point set difference of the given geometries.
ST_Envelope(Geometry) → Geometry	Returns the bounding rectangular polygon of a geometry.
ST_ExteriorRing(Geometry) → Geometry	Returns a line string representing the exterior ring of the input polygon.
ST_Intersection(Geometry, Geometry) → Geometry	Returns the geometry value that represents the point set intersection of two geometries.

Function	Description
ST_SymDifference(Geometry, Geometry) → Geometry	Returns the geometry value that represents the point set symmetric difference of two geometries.

Relationship tests

Function	Description
ST_Contains(Geometry, Geometry) → boolean	Returns true if and only if no points of the second geometry lie in the exterior of the first geometry, and at least one point of the interior of the first geometry lies in the interior of the second geometry. Returns false if points of the second geometry are on the boundary of the first geometry.
ST_Crosses(Geometry, Geometry) → boolean	Returns true if the supplied geometries have some, but not all, interior points in common.
ST_Disjoint(Geometry, Geometry) → boolean	Returns true if the given geometries do not spatially intersect.
ST_Equals(Geometry, Geometry) → boolean	Returns true if the given geometries represent the same geometry.
ST_Intersects(Geometry, Geometry) → boolean	Returns true if the given geometries spatially intersect in two dimensions.
ST_Overlaps(Geometry, Geometry) → boolean	Returns true if the given geometries share space, are of the same dimension, but are not completely contained by each other.
ST_Relate(Geometry, Geometry) → boolean	Returns true if the first geometry is spatially related to the second geometry.
ST_Touches(Geometry, Geometry) → boolean	Returns true if the given geometries have at least one point in common, but their interiors do not intersect.
ST_Within(Geometry, Geometry) → boolean	Returns true if the first geometry is completely inside the second geometry. Returns false if the two geometries have points in common at the boundaries.

Accessors

Function	Description
ST_Area(Geometry) → double	Returns the two-dimensional Euclidean area of a geometry.
ST_Centroid(Geometry) → Geometry	Returns the point value that is the mathematical centroid of a geometry.
ST_CoordDim(Geometry) → bigint	Returns the coordinate dimension of a geometry.
ST_Dimension(Geometry) → bigint	Returns the inherent dimension of a geometry object, which must be less than or equal to the coordinate dimension.
ST_Distance(Geometry, Geometry) → double	Returns the minimum two-dimensional Cartesian distance (based on spatial ref) between two geometries in projected units.
ST_IsClosed(Geometry) → boolean	Returns true if the start and end points of the linestring are coincident.
ST_IsEmpty(Geometry) → boolean	Returns true if Geometry is an empty geometry, such as geometry collection, polygon, and point.
ST_IsRing(Geometry) → boolean	Returns true if and only if the line is closed and simple.
ST_Length(Geometry) → double	Returns the length of a linestring or multi-linestring using Euclidean measurement on a two-dimensional plane (based on spatial ref) in projected units.
ST_XMax(Geometry) → double	Returns X maxima of a bounding box of a geometry.
ST_YMax(Geometry) → double	Returns Y maxima of a bounding box of a geometry.
T_XMin(Geometry) → double	Returns X minima of a bounding box of a geometry.
ST_YMin(Geometry) → double	Returns Y minima of a bounding box of a geometry.

Function	Description
ST_StartPoint(Geometry) → point	Returns the first point of a linestring geometry.
ST_EndPoint(Geometry) → point	Returns the last point of a linestring geometry.
ST_X(Point) → double	Returns the X coordinate of the point.
ST_Y(Point) → double	Returns the Y coordinate of the point.
ST_NumPoints(Geometry) → bigint	Returns the number of points in a geometry.
ST_NumInteriorRing(Geometry) → bigint	Returns the cardinality of the collection of interior rings of a polygon.

18.9.6.27 JOIN syntax

A JOIN operation combines the records of multiple tables. In Log Service, JOIN is applicable to tables in a single Logstore, between Logstore tables and RDS tables, and between tables in different Logstores. This topic describes how to perform a JOIN operation across different Logstores.

Procedure

1. **Download the** *latest version of the Python SDK*.
2. **Call the GetProjectLogs operation to query logs.**

Sample SDK

```
#! /usr/bin/env python
#encoding: utf-8
import time,sys,os
from aliyun.log.logexception import LogException
from aliyun.log.logitem import LogItem
from aliyun.log.logclient import LogClient
from aliyun.log.getlogsrequest import GetLogsRequest
from aliyun.log.getprojectlogsrequest import GetProjectLogsRequest
from aliyun.log.putlogsrequest import PutLogsRequest
from aliyun.log.listtopicsrequest import ListTopicsRequest
from aliyun.log.listlogstoresrequest import ListLogstoresRequest
from aliyun.log.gethistogramsrequest import GetHistogramsRequest
from aliyun.log.index_config import *
from aliyun.log.logtail_config_detail import *
from aliyun.log.machine_group_detail import *
from aliyun.log.acl_config import *
if __name__=='__main__':
    token = None
    endpoint = "http://cn-hangzhou.log.aliyuncs.com"
    accessKeyId  = '******'
    accessKey='*****************'
    client = LogClient(endpoint, accessKeyId, accessKey,token)
    logstore = "meta"
```

```
    # In the query statements, specify two Logstores, the query time
ranges of both Logstores, and the key for Logstore association.
    req = GetProjectLogsRequest(project,"select count(1) from
sls_operation_log  s join  meta m on s.__date__ >'2018-04-10 00:00:00
' and s.__date__ < '2018-04-11 00:00:00' and m.__date__ >'2018-04-23
00:00:00' and m.__date__ <'2018-04-24 00:00:00' and s.projectid = cast
(m.ikey as varchar)");
    res = client.get_project_logs(req)
    res.log_print();
    exit(0)
```

18.9.7 Advanced analysis

18.9.7.1 Optimize queries

The efficiency of an analysis varies across different queries. The following methods are provided to optimize queries:

1. *Avoid using Group By on string columns if possible*
2. *List fields in lexicographical order when you run Group By on multiple columns*
3. *Use estimating functions*
4. *Retrieve required columns by using SQL statements and do not read all columns if possible*
5. *Place non-GROUP BY columns in an aggregate function if possible*

Avoid using Group By on string columns if possible

Using Group By on strings leads to a large number of hash calculations, which account for more than 50% of total calculations.

The following two queries are used in this example:

```
* | select count(1) as pv , date_trunc('hour',__time__) as time group
by time
                * | select count(1) as pv , from_unixtime(__time__-
__time__%3600) as time group by __time__-__time__%3600
```

Both Query 1 and Query 2 calculate the log count every hour. However, Query 1 converts time into a string. For example, `2017-12-12 00:00:00`. Then, Query 1 runs Group By on this string. Query 2 runs Group By on the on-the-hour time value and then converts the result into a string. Query 1 is less efficient than Query 2 because Query 1 needs to hash strings.

List fields in lexicographical order when you run Group By on multiple columns

The following two queries can be used to query data of 13 provinces with 100 million users.

```
Fast: * | select province,uid,count(1) group by province,uid
```

```
                Slow: * | select province,uid,count(1) group by uid,
province
```

Use estimating functions

Estimating functions provide stronger performance than accurate calculation. In estimation, accuracy is compromised to an acceptable extent for fast calculation.

```
Fast: * | select approx_distinct(ip)
                Slow: * | select count(distinct(ip))
```

Retrieve required columns by using SQL statements and do not read all columns if possible

Use the query syntax to retrieve all columns. To speed up SQL calculation, you retrieve only the required columns.

```
Fast: * |select a,b c
                Slow: * |select *
```

Place non-GROUP BY columns in an aggregate function if possible

For example, a user ID must correspond to a username. Therefore, you can only use user IDs to run Group By.

```
Fast: * | select userid, arbitrary(username), count(1)groupby userid
                Slow: * | select userid, username, count(1)groupby
userid,username
```

18.9.7.2 Case study

This topic describes practical cases of log query and analysis.

Case list

1. *Trigger an alert when the error rate exceeds 40% over the last five minutes*
2. *Trigger an alert when traffic plunges*
3. *Calculate the average latency of each bucket set by data interval*
4. *Return percentages included in GROUP BY results*
5. *Count the number of logs that meet the query condition*

Trigger an alert when the error rate exceeds 40% over the last five minutes

Calculate the percentage of error 500 every minute. An alert is triggered when the error rate exceeds 40% over the last five minutes.

```
status:500 | select __topic__, max_by(error_count,window_time)/1.0/sum
(error_count) as error_ratio, sum(error_count) as total_error  from (
                select __topic__, count(*) as error_count , __time__
 - __time__ % 300 as window_time from log group by __topic__,
window_time
```

```
                )
                group by __topic__  having  max_by(error_count,
window_time)/1.0/sum(error_count)   > 0.4  and sum(error_count) > 500
order by total_error desc limit 100
```

Trigger an alert when traffic plunges

Calculate the traffic every minute. An alert is triggered when traffic plunges. Data is collected in less than one minute. Therefore, you must divide the statistical value by (max(time) - min(time)) for normalization to get the average traffic per minute.

```
* | SELECT SUM(inflow) / (max(__time__) - min(__time__))  as
inflow_per_minute, date_trunc('minute',__time__)  as minute group by
minute
```

Calculate the average latency of each bucket set by data interval

```
* | select avg(latency) as latency , case when originSize < 5000 then
 's1' when originSize < 20000 then 's2' when originSize < 500000 then
 's3' when originSize < 100000000 then 's4' else 's5' end as os group
by os
```

Return percentages included in GROUP BY results

List the count results of different departments and their percentages. This query combines subqueries and window functions. sum(c) over() indicates the sum of values in all rows.

```
* |  select   department, c*1.0/ sum(c) over ()  from(select  count(1
) as c, department   from log groupby department)
```

Count the number of logs that meet the query condition

To count the number of URLs based on the different characteristics of URLs, you can use the CASE WHEN syntax. You can also use the count_if syntax, which is simpler.

```
* | select  count_if(uri like '%login') as login_num, count_if(uri
like '%register') as register_num, date_format(date_trunc('minute',
```

```
__time__), '%m-%d %H:%i') as time  group by time order by time limit
100
```

18.9.8 Log analysis through JDBC

In addition to the RESTful API, you can also use the JDBC API and standard SQL-92 statements to query and analyze logs.

Connection parameters

Connection parameter	Example	Description
host	regionid.example.com	The endpoint used to access Log Service.
port	10005	The default port is 10005.
user	bq2sjzesjmo86kq	The AccessKey ID.
password	4fdO1fTDDuZP	The AccessKey Secret.
database	sample-project	The project that is created under the account.
table	sample-logstore	The Logstore that is created for the project .

The following example shows how to use a MySQL command to connect to Log Service:

```
mysql -hcn-shanghai-intranet.log.aliyuncs.com -ubq2sjzesjmo86kq -
p4fdO1fTDDuZP -P10005
use sample-project; // Use a project.
```

Prerequisites

You must use the AccessKey of the main account or a sub-account to use the JDBC API. The sub-account must belong to the project owner and have project-level read permissions.

Syntax

Precautions

The WHERE clauses must contain `__date__` or `__time__` to limit the time range of query. The type of `__date__` is timestamp, and the type of `__time__` is bigint.

Examples:

- `__date__ > '2017-08-07 00:00:00' and __date__ < '2017-08-08 00:00:00'`

- `__time__ > 1502691923 and __time__ < 1502692923`

At least one of the preceding conditions must be met.

Filtering syntax

The filtering syntax in WHERE clauses is as follows:

Meaning	Example	Description
String search	`key = "value"`	Results after word segmentation are queried.
String fuzzy matching	`key = "valu*"`	Results of fuzzy match after word segmentation are queried.
Value comparison	`num_field > 1`	The supported comparison operators include `>`, `>=`, `=`, `<` and `<=`.
Logical operation	`and or not`	For example, `a = "x" and b ="y"` or `a = "x" and not b ="y"`.
Full-text search	`__line__ ="abc"`	Full-text index search requires the special key (`__line__`).

Calculation syntax

For more information about the supported operators, see *Analysis syntax and functions*.

SQL-92 syntax

The SQL-92 syntax is a combination of filtering and calculation syntax.

The following query is used as an example:

```
status>200 |select avg(latency),max(latency) ,count(1) as c GROUP BY
method  ORDER BY c DESC  LIMIT 20
```

The filtering part and time condition in the query can be combined into a new query based on standard SQL-92 syntax.

```
select avg(latency),max(latency) ,count(1) as c from sample-logstore
 where status>200 and __time__>=1500975424 and __time__ < 1501035044
GROUP BY  method  ORDER BY c DESC  LIMIT 20
```

Access Log Service through JDBC

Calls by using programs

Developers can use the MySQL syntax to connect to Log Service in any program that supports MySQL connectors. For example, JDBC or Python MySQLdb can be used.

Example:

```
import com.mysql.jdbc.*;
import java.sql.*;
import java.sql.Connection;
import java.sql.ResultSetMetaData;
import java.sql.Statement;
public class testjdbc {
    public static void main(String args[]){
        Connection conn = null;
        Statement stmt = null;
        try {
            //STEP 2: Register JDBC driver
            Class.forName("com.mysql.jdbc.Driver");
            //STEP 3: Open a connection
            System.out.println("Connecting to a selected database
...");
            conn = DriverManager.getConnection("jdbc:mysql://cn-
shanghai-intranet.log.aliyuncs.com:10005/sample-project","accessid","
accesskey");
            System.out.println("Connected database successfully...") ;
            //STEP 4: Execute a query
            System.out.println("Creating statement...");
            stmt = conn.createStatement();
            String sql = "SELECT method,min(latency,10)  as c,max
(latency,10) from sample-logstore where  __time__>=1500975424 and
__time__ < 1501035044 and latency > 0  and latency < 6142629 and  not
 (method='Postlogstorelogs' or method='GetLogtailConfig') group by
method " ;
            String sql-example2 = "select count(1) ,max(latency),
avg(latency), histogram(method),histogram(source),histogram(status),
histogram(clientip),histogram(__source__) from  test10 where __date__
  >'2017-07-20 00:00:00'  and  __date__ <'2017-08-02 00:00:00' and
__line__='abc#def' and latency < 100000 and (method = 'getlogstorelogS
' or method='Get**' and method <> 'GetCursorOrData' )";
            String sql-example3 = "select count(1) from  sample-
logstore where     __date__  >       '2017-08-07 00:00:00' and
__date__ <     '2017-08-08 00:00:00' limit 100";
            ResultSet rs = stmt.executeQuery(sql);
```

```
            //STEP 5: Extract data from result set
            while(rs.next()){
                //Retrieve by column name
                ResultSetMetaData data = rs.getMetaData();
                System.out.println(data.getColumnCount());
                for(int i = 0;i < data.getColumnCount();++i) {
                    String name = data.getColumnName(i+1);
                    System.out.print(name+":");
                    System.out.print(rs.getObject(name));
                }
                System.out.println();
            }
            rs.close();
        } catch (ClassNotFoundException e) {
            e.printStackTrace();
        } catch (SQLException e) {
            e.printStackTrace();
        } catch (Exception e) {
            e.printStackTrace();
        } finally {
            if (stmt ! = null) {
                try {
                    stmt.close();
                } catch (SQLException e) {
                    e.printStackTrace();
                }
            }
            if (conn ! = null) {
                try {
                    conn.close();
                } catch (SQLException e) {
                    e.printStackTrace();
                }
            }
        }
    }
}
```

Calls by using tools

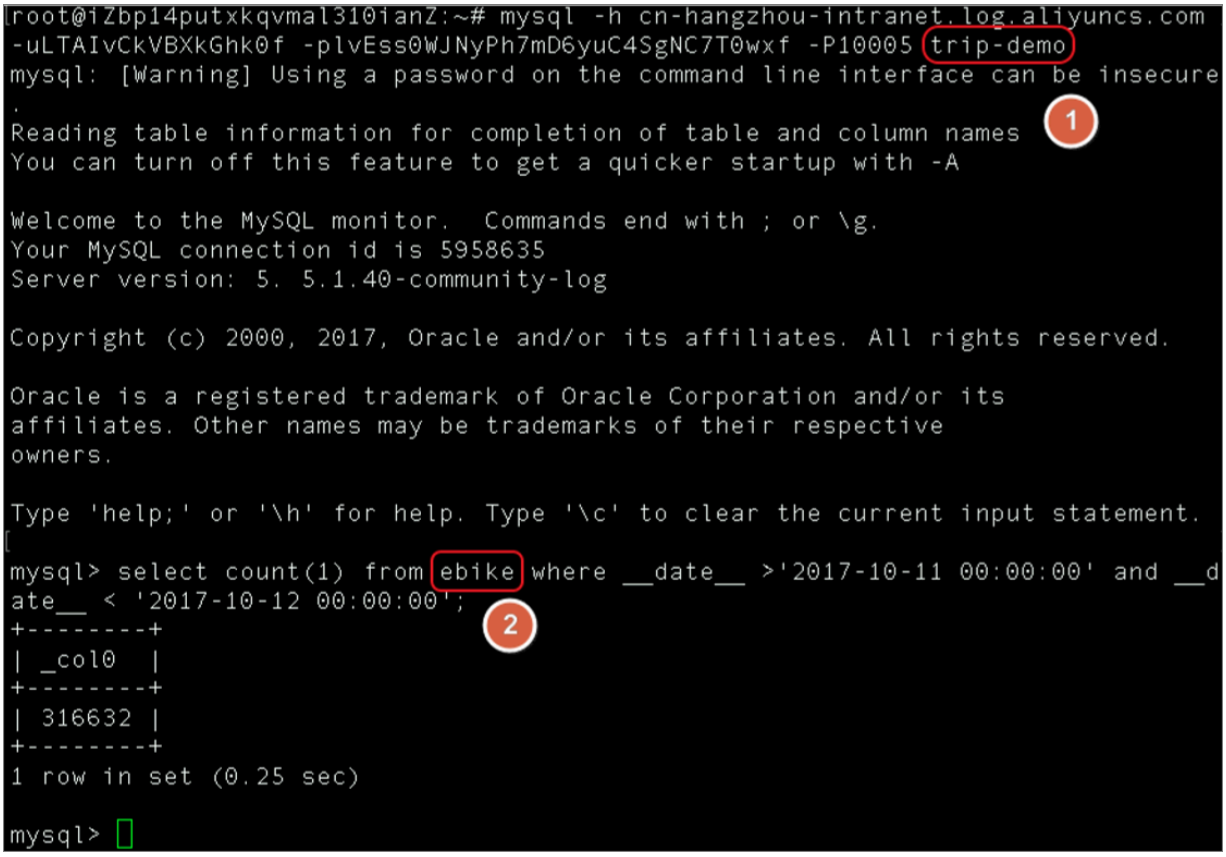
As shown in *Figure 18-3: Connection example*, in a classic network or a VPC, you can use the MySQL client to connect to Log Service.

Note:

1. **Enter your project name at ①.**

2. **Enter your Logstore name at ②.**

Figure 18-3: Connection example



18.10 Alerts

18.10.1 Overview

Log Service enables you to configure alerts based on Saved Search to monitor service status in real time.

Implementation

Log Service implements the alerting feature based on Saved Search. Log on to the Log Service console and go to the query page. Then, set an alert rule and specify the attribute, check condition, and alert action for rule on the page. After an alert rule is configured, Log Service periodically queries the collected log data and sends alert notifications when the query results meet the predefined condition, implementing real-time monitoring of the service status.

- **Alert rule attribute: Set the time and interval of data checks so that Log Service can check as scheduled.**

- **Check condition: Set the comparison field, comparison operator, and check threshold. The comparison operator and the check threshold comprise the check condition. An alert notification is sent when the query results of the comparison field meet the check condition.**
- **Alert action: Set the notification method and alert content. We recommend that you configure notification center as the notification method. If you use this method, alerts are sent to the contacts specified in the notification center through website notifications, text messages, and emails.**

Saved Search in alerts

Saved Search supports query statements or query and analysis statements.

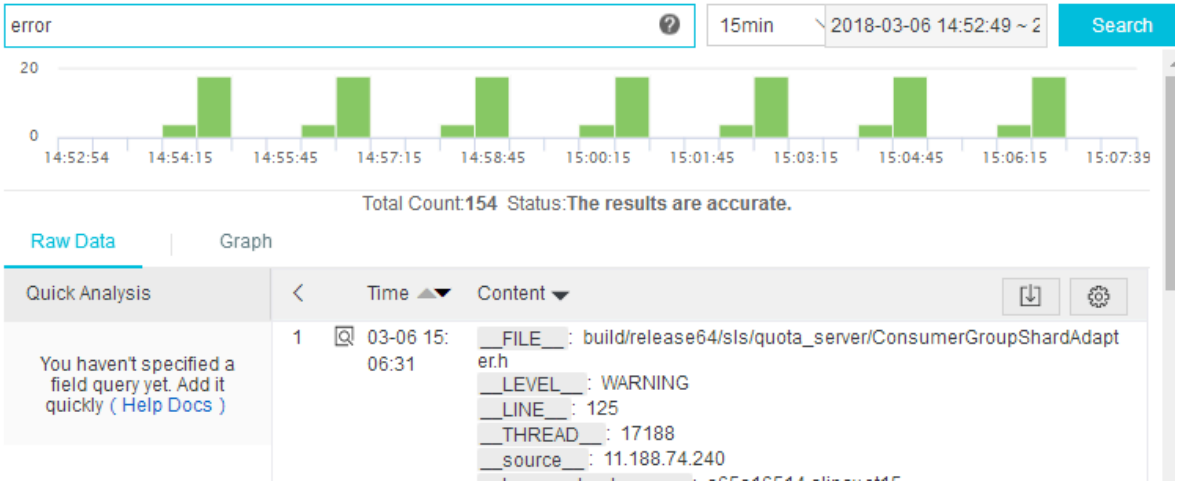
- **Query statement: Log data that matches the query condition is returned.**
- **Query and analysis statement: Statistical analysis is performed on logs that match the query condition and the results of statistical analysis are returned.**
- **Query statement**

  For example, you want to query the data that contains errors in the last 15 minutes, where the condition is error. A total of 154 records are found. The content of each record contains key-value pairs. You can set an alert condition for the corresponding value of a key.

   **Note:**

If the number of query results exceeds 10 in a single query, only the first 10 results are judged by the alarm rules. An alert is triggered when any of the first 10 results meets the condition.

Figure 18-4: Query statement



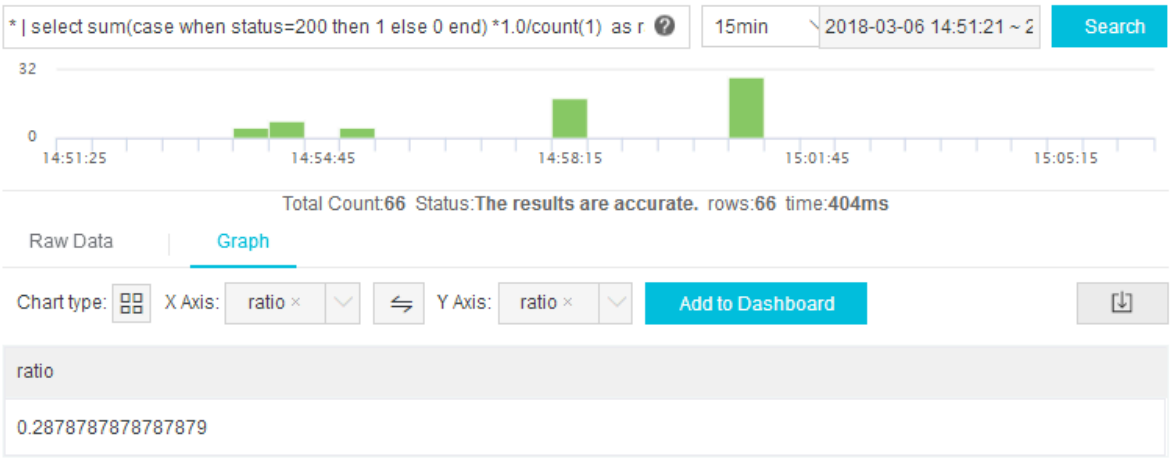
- **Query and analysis statement**

  For example, the following statement is used to query the ratio of logs with the status code 200 in all logs:

```
* | select sum(case when status=200 then 1 else 0 end) *1.0/count(1)
  as ratio
```

  If you set the alert condition as `ratio < 0.9`, an alert is triggered when the ratio of logs with the status code 200 in all logs is less than 90%.

Figure 18-5: Query and analysis statement



18.10.2 Configure an alert

You can save a saved search as an alert on the query page to receive alerts when alert conditions are met.

Prerequisites

- **Log data has been collected.**
- **An index has been configured.**

Procedure

1. *Log on to the Log Service console*.
2. **On the Logstores page, click Search in the LogSearch column.**
3. **Enter a query statement in the search box.**
4. **Set a time range to be queried and click Search.**
5. **Click Save as Alert in the upper-right corner of the page.**
6. **Configure alert rules and click OK.**

Rule	Description
Alert rule name	The name of the alert rule. The name must be 3 to 63 characters in length.
Alert rule attributes	
Saved search name	The name of the saved search used to configure the alert. You can set the name as follows: · Current query, which indicates the current query statement is used as a saved search. · Other existing saved searches.
Data query time (minutes)	The time range of the data read by the server for each alert check. For example, if the value is set to 1, the server will query the data that is generated within one minute before an alert check. You can specify any integer from 1 to 60. **Note:** The server only processes the first 10 data records generated during the time range for sampling purpose when performing an alert check.

Rule	Description
Check interval (minutes)	The interval at which the server performs an alert check. You can specify any integer from 1 to 1440.
Trigger count	The number of consecutive alarm check triggers. An alert notification is sent when the specified trigger count is reached. You can specify any integer from 1 to 10000. For example, if the check interval (minutes) is set to 1 and the trigger count is set to 2, the server performs a check every minute and sends an alert if the results of two consecutive checks meet the alert conditions. The minimum interval between alert notifications is two minutes.
Check condition	
Key name	The name of the key used for alerting in the log.
Comparison operator	The comparison operator of the check condition. It can be of the numeric or character type. For more information about comparison operators, see *Table 18-16: Comparison operator*.
Check threshold	The comparative value in the check condition. This value is combined with the comparison operator to determine whether saved search results meet the alert conditions.
Alert action	
Notification type	The method for sending alert notifications. When the configured alert rule is triggered, Log Service sends an alert based on the preset notification method. Alerts are only sent through the WebHook-Custom method. Notifications are sent to the custom webhook link through the Post method.

Rule	Description
WebHook address	The URL of webhook. The webhook address can contain a maximum of 256 characters. It must start with http:// or https://.
Notification content	The content of an alert notification. The content can contain a maximum of 500 characters.

Table 18-16: Comparison operator

Operator	Description	Example
>	The column value is greater than a value.	$count > 0
<	The column value is smaller than a value.	$count<200
>=	The column value is greater than or equal to a value.	$count>=0
<=	The column value is smaller than or equal to a value.	$count<=0
like	A matching substring.	$project like “admin”
regex	A string that matches the regular expression.	$project regex match “^/S+$”

Result

You can view the detailed alert records after creating alert rules.

1. On the Logstores page, choose Search/Analytics > Alert from the left-side navigation pane.
2. Select an alert rule and click View to view the detailed alert records or the alert status.

Alert status:

- Success: indicates that the rule is executed. You can click Trigger Details to view the trigger conditions.

- **Failure: indicates that the rule fails to be executed during the query, alert rule matching, or notification phase. You can click Trigger Details for more information about the execution failure.**
  - **Query failed: The query syntax is incorrect.**
  - **Query call failed: Check your network connectivity.**
  - **Failed to call the rule: Check whether rule parameters and response data are in the same format.**

18.10.3 Notification methods

You can use webhook custom methods to send alert notifications.

An alert notification contains the following items:

Item	Description
Uid	The ID of an Apsara Stack tenant account (AliUid).
Project	The name of the project for which an alert is generated.
Trigger	The name of the alert.
Condition	The check condition configured for the alert.
Message	The content of the alert notification.
Context	The query results.

WebHook-Custom method

The alert notification method can be set to WebHook-Custom. When an alert is triggered, the alert notification is sent to the custom webhook address by using the Post method.

Procedure

1. **Log on to the Log Service console and set an alert.**
2. **Set Notification Type to WebHook-Custom.**

3. **Enter a custom webhook address in WebHook Address, and then enter the notification content.**

   When an alert is triggered, the alert notification is sent to the custom webhook address by using the Post method.

   Example:

```
{"uid":
            "13415134513","project":"ali-cn","trigger":"oplog_alert
","condition":"3413 >
            3000", "message":"PV count down 30%",
      "context":"c:3413"}
```

18.11 Real-time subscription and consumption

18.11.1 Preview logs

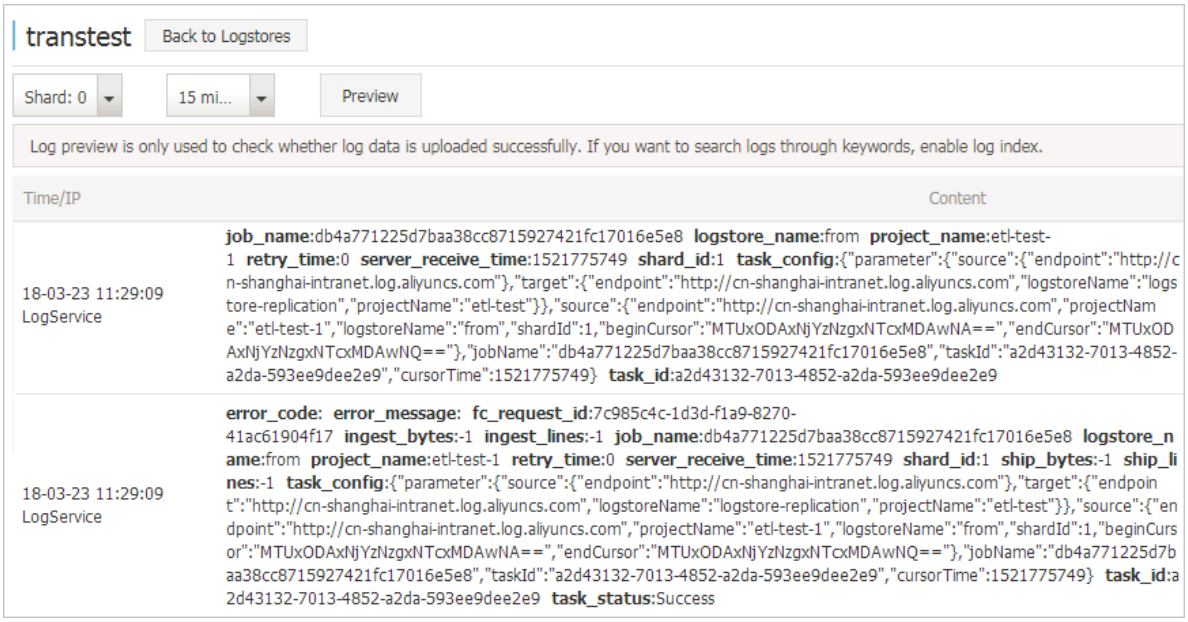
Log preview is a regular type of log consumption. The Log Service console provides a dedicated preview page for you to preview logs in the Logstore by using your browser.

Procedure

1. *Log on to the Log Service console*.
2. **Select a project and click the name of the project to go to the Logstores page.**
3. **On the Logstores page, select a Logstore and click Preview in the Log Consumption column.**

4. **On the log preview page, select a shard of the Logstore and a time range, and click Preview.**

   The log preview page displays the log data of the first 10 packets in the specified time range.



18.11.2 Consumption by consumer groups

18.11.2.1 Consumption by a consumer group

Log Service allows log consumption by consumer group.

A consumer library is an advanced mode of log consumption in Log Service, and introduces the consumer group concept to abstract and manage consumers. Compared with using SDKs to read data, a consumer library can help you focus on the business logic without concerning about implementation details of Log Service. You do not need to worry about load balancing and failover between consumers either.

Terms

You must understand consumer groups and consumers before using a consumer library.

- **consumer group**

  A consumer group consists of multiple consumers. The consumers in a consumer group consume data in a Logstore, and the data consumed by each consumer is different.

- **consumer**

  A consumer is the basic unit to compose a consumer group. Consumers consumes data. The names of consumers in a consumer group must be unique.

A Logstore has multiple shards. A consumer library allocates shards to consumers in a consumer group based on the following principles:

- **Each shard can be allocated to only one consumer.**
- **A consumer can have multiple shards.**

After a new consumer joins a consumer group, affiliations of the shards for the consumer group are adjusted for load balancing. However, the allocation principles remain unchanged, and the allocation process is transparent.

A consumer library can also store checkpoints. This enables consumers to resume consuming data from a breakpoint after a program fault is resolved and makes sure that data is consumed only once.

Instructions

Maven dependencies

```
<dependency>
  <groupId>com.google.protobuf</groupId>
  <artifactId>protobuf-java</artifactId>
  <version>2.5.0</version>
</dependency>
<dependency>
  <groupId>com.aliyun.openservices</groupId>
  <artifactId>aliyun-log</artifactId>
  <version>0.6.11</version>
</dependency>
<dependency>
<groupId>com.aliyun.openservices</groupId>
<artifactId>loghub-client-lib</artifactId>
<version>0.6.15</version>
</dependency>
```

main .java file

```
public class Main {
  // Specify the endpoint of Log Service based on the actual situation
.
  private static String sEndpoint = "cn-hangzhou.log.aliyuncs.com";
  // Specify the project name of Log Service based on the actual
situation.
  private static String sProject = "ali-cn-hangzhou-sls-admin";
  // Specify the Logstore name of Log Service based on the actual
situation.
  private static String sLogstore = "sls_operation_log";
  // Specify the consumer group name based on the actual situation.
  private static String sConsumerGroup = "consumerGroupX";
```

```
   // Specify the AccessKey for data consumption based on the actual
situation.
  private static String sAccessKeyId = "";
  private static String sAccessKey = "";
  public static void main(String []args) throws LogHubClientWorkerEx
ception, InterruptedException
  {
        // The second parameter is the consumer name. The names of
 consumers in a consumer group must be unique. However, the names
of consumer groups can be the same. Different consumer names start
multiple processes on multiple machines to consume a Logstore evenly
. In this case, the names of consumer groups can be differentiated by
machine IP address. The ninth parameter maxFetchLogGroupSize indicates
 the number of log groups retrieved from the server at a time. You can
 use the default value, or adjust the value in the range of (0,1000]
as needed.
      LogHubConfig config = new LogHubConfig(sConsumerGroup, "
consumer_1", sEndpoint, sProject, sLogstore, sAccessKeyId, sAccessKey
, LogHubConfig.ConsumePosition.BEGIN_CURSOR);
      ClientWorker worker = new ClientWorker(new SampleLogHubProcesso
rFactory(), config);
        Thread thread = new Thread(worker);
        // The ClientWorker instance runs automatically after the
thread is executed and extends the Runnable interface.
       thread.start();
       Thread.sleep(60 * 60 * 1000);
       // The shutdown function of the ClientWorker instance is called
 to exit the consumption instance. The associated thread is stopped
automatically.
       worker.shutdown();
       // Multiple asynchronous tasks are generated when the
ClientWorker instance is running. We recommend that you wait 30
seconds so that all running tasks exit after shutdown.
       Thread.sleep(30 * 1000);
  }
}
```

SampleLogHubProcessor.java files

```
public class SampleLogHubProcessor implements ILogHubProcessor
{
  private int mShardId;
  // Record the last persistent checkpoint time.
  private long mLastCheckTime = 0;
  public void initialize(int shardId)
  {
      mShardId = shardId;
  }
  // The main logic of data consumption. All exceptions must be
captured and cannot be thrown.
  public String process(List<LogGroupData> logGroups,
          ILogHubCheckPointTracker checkPointTracker)
  {
      // Display the retrieved data.
      for(LogGroupData logGroup: logGroups){
          FastLogGroup flg = logGroup.GetFastLogGroup();
          System.out.println(String.format("\tcategory\t:\t%s\n\
tsource\t:\t%s\n\ttopic\t:\t%s\n\tmachineUUID\t:\t%s",
                  flg.getCategory(), flg.getSource(), flg.getTopic(),
flg.getMachineUUID()));
          System.out.println("Tags");
          for (int tagIdx = 0; tagIdx < flg.getLogTagsCount(); ++
tagIdx) {
```

```
                FastLogTag logtag = flg.getLogTags(tagIdx);
                System.out.println(String.format("\t%s\t:\t%s", logtag.
getKey(), logtag.getValue()));
            }
            for (int lIdx = 0; lIdx < flg.getLogsCount(); ++lIdx) {
                FastLog log = flg.getLogs(lIdx);
                System.out.println("--------\nLog: " + lIdx + ", time: "
 + log.getTime() + ", GetContentCount: " + log.getContentsCount());
                for (int cIdx = 0; cIdx < log.getContentsCount(); ++cIdx
) {
                    FastLogContent content = log.getContents(cIdx);
                    System.out.println(content.getKey() + "\t:\t" +
content.getValue());
                }
            }
        }
        long curTime = System.currentTimeMillis();
        // Write checkpoints to the server every 30 seconds. If a
ClientWorker instance crashes within 30 seconds,
        // a new ClientWorker instance consumes data starting from the
last checkpoint. Duplicate data may exist.
        if (curTime - mLastCheckTime >  30 * 1000)
        {
            try
            {
                // If the parameter is set to true, checkpoints are
updated to the server immediately. If the parameter is set to false
, checkpoints are cached locally. The default update interval of
checkpoints is 60 seconds.
                checkPointTracker.saveCheckPoint(true);
            }
            catch (LogHubCheckPointException e)
            {
                e.printStackTrace();
            }
            mLastCheckTime = curTime;
        }
        return null;
    }
    // The ClientWorker instance calls this function upon exit. You can
perform a cleanup.
    public void shutdown(ILogHubCheckPointTracker checkPointTracker)
    {
        // Save consumption breakpoints to the server.
        try {
            checkPointTracker.saveCheckPoint(true);
        } catch (LogHubCheckPointException e) {
            e.printStackTrace();
        }
    }
}
class SampleLogHubProcessorFactory implements ILogHubProcessorFactory
{
    public ILogHubProcessor generatorProcessor()
    {
        // Generate a consumption instance.
        return new SampleLogHubProcessor();
    }
}
```

Run the preceding code to print all data in a Logstore. If you need multiple consumers to consume the same Logstore, you can modify the program based on

the comments. You can use the same consumer group name and different consumer names to start a new consumption process.

Limits and troubleshooting

A maximum of 10 consumer groups can be created for each Logstore. The `ConsumerGroupQuotaExceed` **error is reported when the number of consumer groups exceeds 10.**

We recommend that you configure Log4j for the consumer program to throw error messages within consumer groups for troubleshooting. If you save the log4j.properties file to the resources directory and execute the program, the following exception occurs:

```
[WARN ] 2018-03-14 12:01:52,747 method:com.aliyun.openservices.loghub.
client.LogHubConsumer.sampleLogError(LogHubConsumer.java:159)
com.aliyun.openservices.log.exception.LogException: Invalid loggroup
count, (0,1000]
```

The following example shows you how to configure a log4j.properties file:

```
log4j.rootLogger = info,stdout
log4j.appender.stdout = org.apache.log4j.ConsoleAppender
log4j.appender.stdout.Target = System.out
log4j.appender.stdout.layout = org.apache.log4j.PatternLayout
log4j.appender.stdout.layout.ConversionPattern = [%-5p] %d{yyyy-MM-dd
HH:mm:ss,SSS} method:%l%n%m%n
```

18.11.2.2 Consumer group status

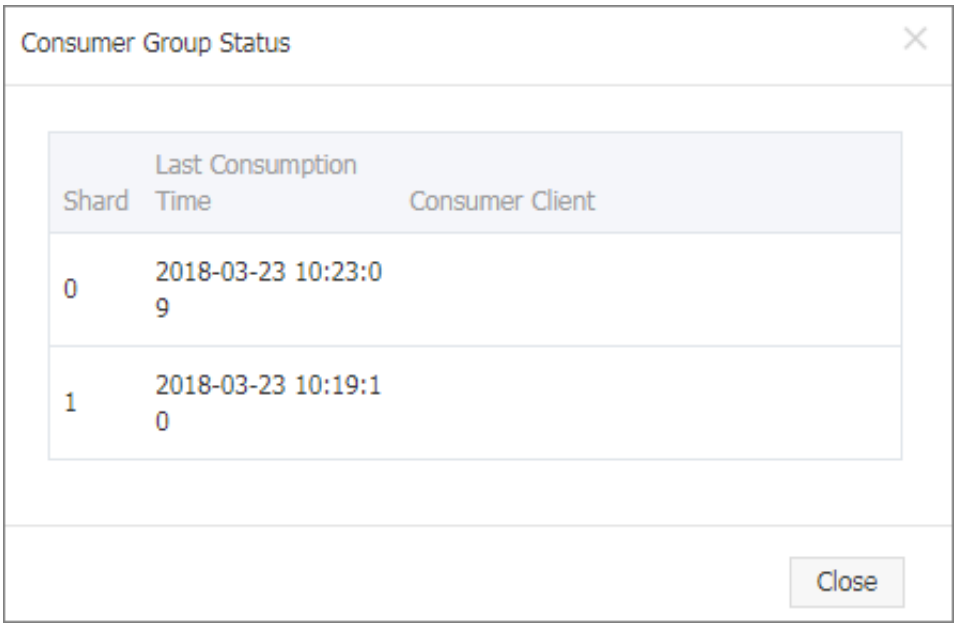
Consumption by a consumer group **is an advanced mode of real-time data consumption. It provides multiple consumption instances for the automatic load balancing of Logstore consumption. Both Spark Streaming and Storm use consumer groups as the basic mode.**

Procedure

1. *Log on to the Log Service console*.
2. **Select a project and click the project name.**
3. **In the left-side navigation pane, choose LogHub - Real-Time Consumption > Consumer Group Management.**

4. On the Consumer Groups page, select a Logstore to check whether collaborative consumption is enabled.

5. Select the specified consumer group and click Status to view the data consumption progress for each shard.



As shown in the preceding figure, the Logstore has four shards that are consumed by four consumers. The last consumption time for each consumer is shown in the second column. You can use the data consumption time to determine whether the current data processing capacity can keep up with data generation. If data processing lags behind, or consumption is slower than data generation, we recommend that you increase the number of consumers.

The Java SDK is used as an example to describe how to call API operations to view consumption status.

```
package test;
import java.util.ArrayList;
import com.aliyun.openservices.log.Client;
import com.aliyun.openservices.log.common.Consts.CursorMode;
import com.aliyun.openservices.log.common.ConsumerGroup;
import com.aliyun.openservices.log.common.ConsumerGr
oupShardCheckPoint;
import com.aliyun.openservices.log.exception.LogException;
public class ConsumerGroupTest {
```

```
    static String endpoint = "";
    static String project = "";
    static String logstore = "";
    static String accesskeyId = "";
    static String accesskey = "";
    public static void main(String[] args) throws LogExcepti
on {
        Client client = new Client(endpoint, accesskeyId,
accesskey);
        // Obtain all consumer groups in this Logstore. If
no consumer group exists, the length of consumerGroups is 0.
        ArrayList<ConsumerGroup>  consumerGroups;
        try{
            consumerGroups = client.ListConsumerGroup(
project, logstore).GetConsumerGroups();
        }
        catch(LogException e){
            if(e.GetErrorCode() == "LogStoreNotExist")
                System.out.println("this logstore does not
have any consumer group");
            else{
                //internal server error branch
            }
            return;
        }
        for(ConsumerGroup c: consumerGroups){
            // Display consumer group properties, including
names, heartbeat timeout, and whether to consume in order.
            System.out.println("Name:" + c.getConsume
rGroupName());
            System.out.println("Heartbeat timeout:" + c.
getTimeout());
            System.out.println("Consumption in order" + c.
isInOrder());
            for(ConsumerGroupShardCheckPoint cp: client.
GetCheckPoint(project, logstore, c.getConsumerGroupName()).
GetCheckPoints()){
                System.out.println("shard: " + cp.getShard
());
                // Format the returned time. The time is a
long integer and accurate to milliseconds.
                System.out.println("The last time when data
is consumed : " + cp.getUpdateTime());
                System.out.println("Consumer name: " + cp.
getConsumer());
                String consumerPrg = "";
                if(cp.getCheckPoint().isEmpty())
                    consumerPrg = "Consumption not started";
                else{
                     // The UNIX timestamp, in seconds.
Format the output value of the timestamp.
                    try{
                        int prg = client.GetPrevCursorTime
(project, logstore, cp.getShard(), cp.getCheckPoint()).
GetCursorTime();
                        consumerPrg = "" + prg;
                    }
                    catch(LogException e){
                        if(e.GetErrorCode() == "InvalidCur
sor")
                            consumerPrg = "Invalid. The
previous consumption time has exceeded the data lifecycle in
 the Logstore.";
                        else{
```

```
                                //internal server error
                                throw e;
                            }
                        }
                    }
                    System.out.println("Consumption progress:"
 + consumerPrg);
                    String endCursor = client.GetCursor(project
, logstore, cp.getShard(), CursorMode.END).GetCursor();
                    int endPrg = 0;
                    try{
                        endPrg = client.GetPrevCursorTime(
project, logstore, cp.getShard(), endCursor).GetCursorTime
();
                    }
                    catch(LogException e){
                        //do nothing
                    }
                     // The UNIX timestamp, in seconds. Format
the output value of the timestamp.
                    System.out.println("The arrival time of the
last piece of data:" + endPrg);
                }
            }
        }
}
```

18.11.3 Use Flink to consume data

Fink log connector is a tool provided by Log Service and used to connect to Flink.

It consists of Flink log consumer and Flink log producer.

The Flink log consumer reads data from Log Service. It supports exactly-once semantics and shard-based load balancing.

The Flink log producer writes data into Log Service. When using the Fink log connector, you must add Maven dependencies to projects. You can use the following code snippets to add Maven dependencies to projects:

```
<dependency>
            <groupId>org.apache.flink</groupId>
            <artifactId>flink-streaming-java_2.11</artifactId>
            <version>1.3.2</version>
            </dependency>
            <dependency>
            <groupId>com.aliyun.openservices</groupId>
            <artifactId>flink-log-connector</artifactId>
            <version>0.1.7</version>
            </dependency>
            <dependency>
            <groupId>com.google.protobuf</groupId>
            <artifactId>protobuf-java</artifactId>
            <version>2.5.0</version>
            </dependency>
            <dependency>
            <groupId>com.aliyun.openservices</groupId>
            <artifactId>aliyun-log</artifactId>
```

```
            <version>0.6.10</version>
            </dependency>
            <dependency>
            <groupId>com.aliyun.openservices</groupId>
            <artifactId>log-loghub-producer</artifactId>
            <version>0.1.8</version>
            </dependency>
```

Prerequisites

1. *Log on to the Log Service console*.
2. **You have an AccessKey pair, and have created a project and a Logstore.**

Log consumer

The Flink log consumer provides the capability of subscribing to a specific Logstore in Log Service to achieve the exactly-once semantics. The Flink log consumer automatically detects the change of the number of shards in a Logstore, making it convenient for the user.

Each Flink sub-job consumes some shards in a Logstore. If shards in a Logstore are split or merged, the shards consumed by sub-jobs will change accordingly.

Associated API operations

The Flink log consumer uses the following Log Service API operations:

- **GetCursorOrData**

  You can call this operation to pull data from a shard. Calling this operation frequently may exhaust the shard quota of Log Service. You can use ConfigConstants.LOG_FETCH_DATA_INTERVAL_MILLIS and ConfigConstants.LOG_MAX_NUMBER_PER_FETCH to control the interval of API calls and number of logs pulled by each call. For more information about the shard quota, see *Split shards*.

  ```
  configProps.put(ConfigConstants.LOG_FETCH_DATA_INTERVAL_MILLIS,  "
  100");
                          configProps.put(ConfigConstants.LOG_MAX_NU
  MBER_PER_FETCH,  "100");
  ```

- **ListShards**

  You can call this operation to view all shards in a Logstore and the status of each shard. If the shards are frequently split and merged, you can adjust the call interval to detect the changes of the shards.

  ```
  // Call the ListShards operation once every 30 seconds
  ```

```
                        configProps.put(ConfigConstants.LOG_SHARDS
_DISCOVERY_INTERVAL_MILLIS,  "30000");
```

- **CreateConsumerGroup**

  You can call this operation to create a consumer group to synchronize checkpoints. This operation can be called only when consumption progress monitoring is enabled.

- **ConsumerGroupUpdateCheckPoint**

  You can call this operation to synchronize snapshots of Flink to a consumer group.

Procedure

1. Configure startup parameters

```
Properties configProps = new Properties();
                    // Set the endpoint to access Log Service.
                    configProps.put(ConfigConstants.LOG_ENDPOINT,  "cn
-hangzhou.log.aliyuncs.com");
                    // Set the AccessKey.
                    configProps.put(ConfigConstants.LOG_ACCESSSKEYID,
 "");
                    configProps.put(ConfigConstants.LOG_ACCESSKEY,
 "");
                    // Set the project.
                    configProps.put(ConfigConstants.LOG_PROJECT,  "ali
-cn-hangzhou-sls-admin");
                    // Set the Logstore.
                    configProps.put(ConfigConstants.LOG_LOGSTORE,  "
sls_consumergroup_log");
                    // Set the start position to consume logs.
                    configProps.put(ConfigConstants.LOG_CONSUM
ER_BEGIN_POSITION,  Consts.LOG_END_CURSOR);
                    // Set the message deserialization method.
                    RawLogGroupListDeserializer deserializer = new
RawLogGroupListDeserializer();
                    final StreamExecutionEnvironment env = StreamExec
utionEnvironment.getExecutionEnvironment();
                    DataStream<RawLogGroupList> logTestStream = env.
addSource(
                    new FlinkLogConsumer<RawLogGroupList>(deserializer
,  configProps));
```

The preceding example describes log consumption. java.util.Properties is used as the configuration tool. The configurations of all consumers are available in ConfigConstants.

Note:

The number of sub-jobs in the Flink stream is independent from that of shards in a Logstore. If the number of shards is greater than that of sub-jobs, each sub-job

consumes multiple shards exactly once. If the number of shards is smaller than that of sub-jobs, some sub-jobs are idle until new shards are generated.

2. Set the consumption start position

The Flink log consumer enables you to set the start position for consuming a shard. By specifying ConfigConstants.LOG_CONSUMER_BEGIN_POSITION, you can start to consume a shard from its header or tail or at a specific point in time. The connector also supports consumption restoration from a specific consumer group. Valid values:

- **Consts.LOG_BEGIN_CURSOR: indicates that you start to consume a shard from its header, which is the earliest data in a shard.**
- **Consts.LOG_END_CURSOR: indicates that you start to consume a shard from its tail, which is the latest data in a shard.**
- **Consts.LOG_FROM_CHECKPOINT: indicates that you start to consume a shard from a checkpoint stored in a specific consumer group. You can specify a consumer group by setting ConfigConstants.LOG_CONSUMERGROUP.**
- **UnixTimestamp: a string of the Integer type. The timestamp is the number of seconds that have elapsed since 00:00:00 Thursday, January 1 1970. It indicates that data in a shard is consumed from this time point.**

The following examples show the valid values:

```
configProps.put(ConfigConstants.LOG_CONSUMER_BEGIN_POSITION, Consts.
LOG_BEGIN_CURSOR);
                configProps.put(ConfigConstants.LOG_CONSUMER_BEGIN_P
OSITION, Consts.LOG_END_CURSOR);
                configProps.put(ConfigConstants.LOG_CONSUMER_BEGIN_P
OSITION, "1512439000");
                configProps.put(ConfigConstants.LOG_CONSUMER_BEGIN_P
OSITION, Consts.LOG_FROM_CHECKPOINT);
```

Note:

If you have configured consumption restoration from state backends of Flink when you start a Flink job, the connector uses checkpoints stored in the state backends.

3. (Optional) Configure consumption progress monitoring

The Flink log consumer enables you to configure consumption progress monitoring . Consumption progress indicates the real-time consumption position of each shard . These positions are expressed by timestamps.

```
configProps.put(ConfigConstants.LOG_CONSUMERGROUP, "your consumer
group name”);
```

Note:

The preceding parameter is optional. If the parameter is specified , the Flink log consumer creates a consumer group first. If a consumer group already exists, no further action is performed. Snapshots in the Flink log consumer are automatically synchronized to the consumer group of Log Service. You can view the consumption progress of the Flink log consumer in the Log Service console.

4. Configure parameters for disaster recovery and exactly-once semantics

If the checkpointing feature of Flink is enabled, the Flink log consumer periodically stores the consumption progress of each shard. When a job fails, Flink restores the consumption progress to Flink log consumer and starts to consume from the latest checkpoint.

The checkpoint interval defines the maximum amount of data to be rolled back, or re-consumed in the event of a failure. You can use the following code:

```
final StreamExecutionEnvironment env = StreamExecutionEnvironment.
getExecutionEnvironment();
                // Enable exactly-once semantics.
                env.getCheckpointConfig().setCheckpointingMode(
CheckpointingMode.EXACTLY_ONCE);
                // Store checkpoints every five seconds.
                env.enableCheckpointing(5000);
```

For more information about the Flink checkpoints, see *Checkpoints* in the Flink documentation.

Log Producer

The Flink log producer writes data into Log Service.

Note:

The Flink log producer supports only the Flink at-least-once semantics. When a job fails, data written into Log Service may be duplicated, but can never be lost.

Procedure

1. **Initialize the Flink log producer.**
   a. **Initialize properties for the Flink log producer.**

      The initialization process for the Flink log producer is similar to that for the Flink log consumer. The Flink log producer contains the following parameters. You can use the default values of these parameters. If necessary, you can customize the values.

```
// The number of I/O threads used to send data. Default value: 8.
                        ConfigConstants.LOG_SENDER_IO_THREAD_COUNT
                        // The time it takes to send the data
after log data is cached. Default value: 3,000.
                        ConfigConstants.LOG_PACKAGE_TIMEOUT_MILLIS
                        // The number of logs in the cached
package. Default value: 4,096.
                        ConfigConstants.LOG_LOGS_COUNT_PER_PACKAGE
                        // The size of the cached package. Default
 value: 3 MB.
                        ConfigConstants.LOG_LOGS_BYTES_PER_PACKAGE
                        // The total memory size that the job can
use. Default value: 100 MB.
                        ConfigConstants.LOG_MEM_POOL_BYTES
```

      These parameters are optional. You can use their default values.

   b. **Reload LogSerializationSchema to define the method of serializing data into RawLogGroup.**

      RawLogGroup is a collection of logs. For more information about the meaning of each field, see the Data model section of the *Log Service Developer Guide***.**

      To use the shard hash key feature of Log Service, you must specify the shard to which data is written. You can use LogPartitioner to generate the hash key for the data.

      Example:

```
FlinkLogProducer<String> logProducer = new FlinkLogProducer<String
>(new SimpleLogSerializer(),  configProps);
                        logProducer.setCustomPartitioner(new
LogPartitioner<String>() {
                        // Generate a 32-bit hash value.
                        public String getHashKey(String element) {
                        try {
                        MessageDigest md = MessageDigest.
getInstance("MD5");
                        md.update(element.getBytes());
                        String hash = new BigInteger(1,  md.digest
()).toString(16);
                        while(hash.length() < 32) hash = "0" +
hash;
```

```
                        return hash;
                        } catch (NoSuchAlgorithmException e) {
                        }
                        return  "00000000000000000000000000000000
00000000000000000000000000000000";
                        }
                        });
```

Note:

LogPartitioner is optional. If this parameter is not specified, data is randomly written into a shard.

2. **Execute the following statements and write the generated string to Log Service.**

```
// Serialize data to the data format of Log Service.
                    class SimpleLogSerializer implements LogSeriali
zationSchema<String> {
                    public RawLogGroup serialize(String element) {
                    RawLogGroup rlg = new RawLogGroup();
                    RawLog rl = new RawLog();
                    rl.setTime((int)(System.currentTimeMillis() /
1000));
                    rl.addContent("message",  element);
                    rlg.addLog(rl);
                    return rlg;
                    }
                    }
                    public class ProducerSample {
                    public static String sEndpoint = "cn-hangzhou.
log.aliyuncs.com";
                    public static String sAccessKeyId = "";
                    public static String sAccessKey = "";
                    public static String sProject = "ali-cn-hangzhou
-sls-admin";
                    public static String sLogstore = "test-flink-
producer";
                    private static final Logger LOG = LoggerFactory.
getLogger(ConsumerSample.class);
                    public static void main(String[] args) throws
Exception {
                    final ParameterTool params = ParameterTool.
fromArgs(args);
                    final StreamExecutionEnvironment env =
StreamExecutionEnvironment.getExecutionEnvironment();
                    env.getConfig().setGlobalJobParameters(params);
                    env.setParallelism(3);
                    DataStream<String> simpleStringStream = env.
addSource(new EventsGenerator());
                    Properties configProps = new Properties();
                    // Set the endpoint to access Log Service.
                    configProps.put(ConfigConstants.LOG_ENDPOINT,
sEndpoint);
                    // Set the AccessKey to access Log Service.
                    configProps.put(ConfigConstants.LOG_ACCESSSKEYID
,  sAccessKeyId);
                    configProps.put(ConfigConstants.LOG_ACCESSKEY,
sAccessKey);
                    // Set the project to which logs are written.
                    configProps.put(ConfigConstants.LOG_PROJECT,
sProject);
```

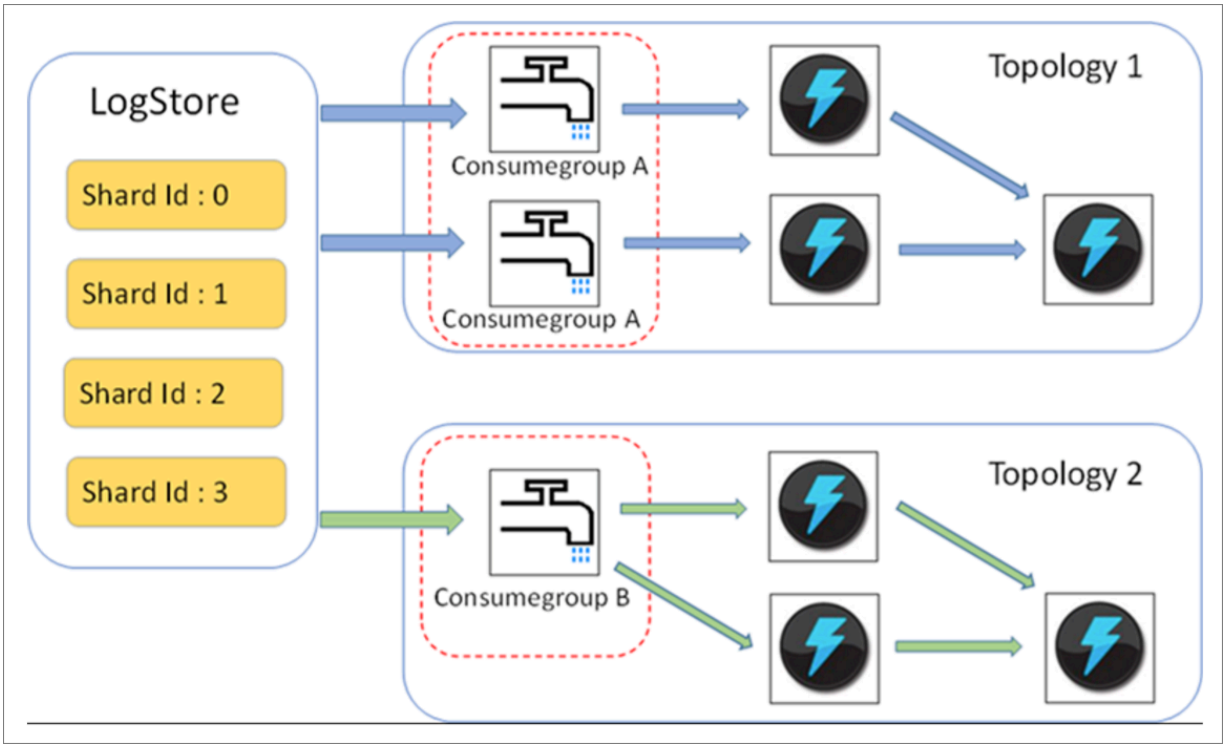
```
                    // Set the Logstore to which logs are written.
                    configProps.put(ConfigConstants.LOG_LOGSTORE,
sLogstore);
                    FlinkLogProducer<String> logProducer = new
FlinkLogProducer<String>(new SimpleLogSerializer(),  configProps);
                    simpleStringStream.addSink(logProducer);
                    env.execute("flink log producer");
                    }
                    // Simulate log generation.
                    public static class EventsGenerator implements
SourceFunction<String> {
                    private boolean running = true;
                    @Override
                    public void run(SourceContext<String> ctx)
throws Exception {
                    long seq = 0;
                    while (running) {
                    Thread.sleep(10);
                    ctx.collect((seq++) + "-" + RandomStringUtils.
randomAlphabetic(12));
                    }
                    }
                    @Override
                    public void cancel() {
                    running = false;
                    }
                    }
                    }
```

18.11.4 Use Storm to consume data

LogHub enables you to collect real-time log data in an efficient and reliable manner by using multiple methods such as Logtail and SDKs. You can access real-time systems such as Spark Streaming and Storm to consume the data that is written to LogHub.

To reduce the cost of LogHub data consumption, Log Service provides Storm users with LogHub Storm spouts to read LogHub data in real time.

Architecture and process



- **In the preceding figure, LogHub Storm spouts are enclosed in the red dotted boxes. Each Storm topology has a group of spouts. Spouts within a group can read all data from a Logstore. Spouts in different topologies are independent of each other.**
- **Each topology is identified by a unique LogHub consumer group name. Spouts in the same topology use LogHub client library to implement load balancing and automatic failover.**
- **Spouts read LogHub data in real time, send the data to bolts in a topology, and then save consumption checkpoints to the LogHub server on a regular basis.**

Limits

- **A maximum of five consumer groups can be created for each Logstore to improve resource utilization. You can call the DeleteConsumerGroup operation of the Java SDK to delete consumer groups that you no longer use.**
- **We recommend that you configure the same number of spouts and shards. Otherwise, a single spout may be unable to process a large amount of data.**
- **If a shard contains a large amount of data, and exceeds the processing capability of a spout, you can call the SplitShard operation to split shards, reducing the amount of data in each shard.**

- **The Storm acknowledgment (ACK) method must be called on LogHub spouts. The ACK method is used to confirm whether spouts have sent messages to bolts correctly. Therefore, bolts must call the ACK method to confirm receipt of the messages.**

Examples

- **Use spouts to create a topology**

```
     public static void main( String[] args )
    {
        String mode = "Local"; // The local test mode.
           String conumser_group_name = ""; // Specify a unique
consumer group name for each topology. The name cannot be an empty
 string. It must be 3 to 63 characters in length and can contain
lowercase letters, digits, hyphens (-), and underscores (_). It must
 start and end with a lowercase letter or digit.
        String project = ""; // The Log Service project.
        String logstore = ""; // The Logstore of Log Service.
        String endpoint = ""; // The endpoint used to access Log
Service.
        String access_id = ""; // The AccessKey of the user.
        String access_key = "";
        // Construct the configurations required for a LogHub Storm
spout.
        LogHubSpoutConfig config = new LogHubSpoutConfig(conumser_g
roup_name,
                endpoint, project, logstore, access_id,
                access_key, LogHubCursorPosition.END_CURSOR);
        TopologyBuilder builder = new TopologyBuilder();
        // Create a LogHub Storm spout.
        LogHubSpout spout = new LogHubSpout(config);
         // In actual use, you can set the number of spouts to the
same number of Logstore shards.
        builder.setSpout("spout", spout, 1);
        builder.setBolt("exclaim", new SampleBolt()).shuffleGrouping
("spout");
        Config conf = new Config();
        conf.setDebug(false);
        conf.setMaxSpoutPending(1);
        // Configure the serialization method LogGroupDataSerializ
Serializer of LogGroupData if Kryo is used to serialize and
deserialize data.
        Config.registerSerialization(conf, LogGroupData.class,
LogGroupDataSerializSerializer.class);
        if (mode.equals("Local")) {
            logger.info("Local mode...") ;
            LocalCluster cluster  = new LocalCluster();
            cluster.submitTopology("test-jstorm-spout", conf,
builder.createTopology());
            try {
                Thread.sleep(6000 * 1000);    //waiting for several
minutes
            } catch (InterruptedException e) {
                // TODO Auto-generated catch block
                e.printStackTrace();
            }
            cluster.killTopology("test-jstorm-spout");
            cluster.shutdown();
        } else if (mode.equals("Remote")) {
```

```
            logger.info("Remote mode...");
            conf.setNumWorkers(2);
            try {
                StormSubmitter.submitTopology("stt-jstorm-spout-4",
conf, builder.createTopology());
            } catch (AlreadyAliveException e) {
                // TODO Auto-generated catch block
                e.printStackTrace();
            } catch (InvalidTopologyException e) {
                // TODO Auto-generated catch block
                e.printStackTrace();
            }
        } else {
            logger.error("invalid mode: " + mode);
        }
    }
}
```

- **Sample code of the bolts that consume data (only the content of each log entry is printed)**

```
public class SampleBolt extends BaseRichBolt {
    private static final long serialVersionUID = 4752656887
774402264L;
    private static final Logger logger = Logger.getLogger(BaseBasicB
olt.class);
    private OutputCollector mCollector;
    @Override
    public void prepare(@SuppressWarnings("rawtypes") Map stormConf
, TopologyContext context,
            OutputCollector collector) {
        mCollector = collector;
    }
    @Override
    public void execute(Tuple tuple) {
        String shardId = (String) tuple
                .getValueByField(LogHubSpout.FIELD_SHARD_ID);
        @SuppressWarnings("unchecked")
        List<LogGroupData> logGroupDatas = (ArrayList<LogGroupData
>) tuple.getValueByField(LogHubSpout.FIELD_LOGGROUPS);
        for (LogGroupData groupData : logGroupDatas) {
            // Each log group consists of one or more logs.
            LogGroup logGroup = groupData.GetLogGroup();
            for (Log log : logGroup.getLogsList()) {
                StringBuilder sb = new StringBuilder();
                // Each log entry has a time field and multiple key-
value pairs.
                int log_time = log.getTime();
                sb.append("LogTime:").append(log_time);
                for (Content content : log.getContentsList()) {
                    sb.append("\t").append(content.getKey()).append
(":")
                            .append(content.getValue());
                }
                logger.info(sb.toString());
            }
        }
        // The Storm acknowledgment (ACK) method must be called on
LogHub spouts. The ACK method is used to confirm whether spouts have
 sent messages to bolts correctly.
        // Therefore, Bolts must call the ACK method to confirm
receipt of the messages.
```

```
            mCollector.ack(tuple);
        }
        @Override
        public void declareOutputFields(OutputFieldsDeclarer declarer) {
            //do nothing
        }
    }
```

Maven

Use the following code to add Maven dependencies for versions earlier than Storm 1 .0 (for example, 0.9.6):

```
<dependency>
  <groupId>com.aliyun.openservices</groupId>
  <artifactId>loghub-storm-spout</artifactId>
  <version>0.6.5</version>
</dependency>
```

Use the following code to add Maven dependencies for Storm 1.0 and later:

```
<dependency>
  <groupId>com.aliyun.openservices</groupId>
  <artifactId>loghub-storm-1.0-spout</artifactId>
  <version>0.1.2</version>
</dependency>
```

18.11.5 Use Spark Streaming to consume data

Log Service enables you to consume logs in real time by using Spark Streaming.

E-MapReduce provides a group of universal APIs for Spark Streaming to consume LogHub data in real time. To download SDKs, go to *GitHub*.

18.11.6 Use StreamCompute to consume data

StreamCompute can be used to consume LogHub data after data sources of the LogHub type are created.

StreamCompute can be used to consume LogHub data after data sources of the LogHub type are created. The following configurations are used:

```
CREATE STREAM TABLE source_test_galaxy ( $schema ) WITH ( type='loghub
',
```

```
endpoint=$endpoint, accessId=$loghub_access_id, accessKey=$loghub_acc
ess_key, projectName=$project, logstore=$logstore );
```

Table 18-17: Parameters

Parameter	Description
`$schema`	**The schema that maps the keys in logs to the columns in the StreamCompute table. Example:** `name STRING, age STRING, id STRING`**.**
`$endpoint`	**Your endpoint.**
`$loghub_access_id`	**The AccessID of the Apsara Stack tenant account (or RAM user) with read permissions.**
`$loghub_access_key`	**The AccessKey of the Apsara Stack tenant account (or RAM user) with read permissions.**
`$project`	**The project where data is located.**
`$logstore`	**The Logstore where data is located.**

Example:

```
CREATE STREAM TABLE source_test_galaxy ( name STRING, age STRING, id
 STRING ) WITH ( type='loghub', endpoint='http://cn-hangzhou-intranet
.log.aliyuncs.com', accessId='mock_access_id', accessKey='mock_acces
s_key', projectName='ali-cloud-streamtest', logstore='stream-test' );
```

19 Apsara Stack Security

19.1 What is Apsara Stack Security?

Apsara Stack Security is a solution that provides Apsara Stack with a full suite of security features, such as network security, server security, application security, data security, and security management.

Background

Traditional security solutions for IT services detect attacks on the network perimeter. They use hardware products such as firewalls and intrusion prevention systems (IPSs) to block attacks outside the network.

With the development of cloud computing that features low costs, flexible configuration on demand, and high resource utilization, an increasing number of enterprises and organizations are switching from traditional IT services to cloud computing services. A cloud computing environment does not have a definite network perimeter. As a result, traditional security solutions cannot effectively protect the security of cloud assets.

Apsara Stack Security combines the powerful data analysis capabilities of Alibaba Cloud with the expertise of the Alibaba Cloud security operations team. It provides integrated security protection services at the network layer, application layer, and server layer.

Complete security solution

Apsara Stack Security consists of Apsara Stack Security Standard Edition and optional security services, and provides users with a complete security solution.

Security domain	Service	Description
Security management	Threat Detection Service	Monitors traffic and overall security status to implement security audit and centralized management.
Server security	Server Guard	Protects Elastic Compute Service (ECS) instances against intrusions and malicious code.

Security domain	Service	Description
	Server Intrusion Detection	Protects physical servers against intrusions.
Application security	Web Application Firewall	Protects Web applications against malicious attacks and guarantees that mobile and PC users from the Internet can securely access Web applications.
Network security	Anti-DDoS	Guarantees the availability of network links and improves business continuity.
Data security	Sensitive Data Discovery and Protection	Protects sensitive data to prevent data leaks and meet compliance requirements.
Security O&M service	On-premises security operations services	Help you fully utilize the security features of Apsara Stack Security and other Apsara Stack services, establish and continuously optimize your cloud security defense system, and guarantee the security of your business systems.

19.2 Restrictions

Before logging on to Apsara Stack Security Center, make sure that your local PC meets the requirements.

Table 19-1: Configuration requirements

Item	Requirements
Browser	· Internet Explorer: 11 or later · Google Chrome (recommended): 42.0.0 or later · Mozilla Firefox: 30 or later · Safari: 9.0.2 or later
Operating system	· Windows XP, Windows 7, or later · Mac

19.3 Quick start

19.3.1 User permissions

This topic describes the user roles involved in Apsara Stack Security.

All roles in Apsara Stack Security Center are default roles. You cannot add custom roles. Before logging on to Apsara Stack Security Center, make sure that your account has been assigned the corresponding role. For more information about roles in Apsara Stack Security, see *Table 19-2: Default roles in Apsara Stack Security*.

Table 19-2: Default roles in Apsara Stack Security

Role	Description
System administrator of Apsara Stack Security Center	Manages and configures the system settings for Apsara Stack Security Center. The system administrator has the following permissions: Alibaba Cloud account management, rule database synchronization, alert settings, and global settings.
Security administrator of Apsara Stack Security Center	Monitors the security status of the entire Apsara Stack platform and configures security policies for each functional module of Apsara Stack Security. The security administrator has permissions to all functional nodes under Threat Detection, Network Security, Application Security, Server Security, PM Security, and Asset Management. **Note:** The permissions to WAF and Cloud Firewall must be assigned independently.

Role	Description
Department security administrator	Monitors the security status of cloud product resources in the specified department and configures security policies for each functional module of Apsara Stack Security for this department. The department security administrator has permissions to all functional nodes under Threat Detection, Network Security, Application Security, Server Security, PM Security, and Asset Management. In addition, the department security administrator can specify the alert notification method and the alert recipients in this department.  **Note:** The permissions to WAF and Cloud Firewall must be assigned independently.
Auditor of Apsara Stack Security Center	Conducts security audits on the entire Apsara Stack platform. The auditor can view audit events and original logs, configure audit policies, and access all functional nodes under Security Audit.

If you do not have an account and a role, contact the administrator to create an account and assign a role to it. For more information, see Create a user in *User Guide*.

19.3.2 Log on to Apsara Stack Security Center

This topic describes how to log on to Apsara Stack Security Center.

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://`IP address or domain name of the Apsara Stack console`/manage.
- We recommend that you use the Chrome browser.

Context

To log on to Apsara Stack Security Center, you must have the permissions of user roles in Security Center. For more information, see *Apsara Stack Security Center User Role Permissions*.

Procedure

1. Open your browser.

2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click LOGIN to go to the Dashboard page.
5. In the top menu bar, click  .
6. In Cloud Security Center, click any service, for example, Server Security.
7. Select a Region, and click Cloud Security Console. The Apsara Stack Security Center page appears.

19.3.3 Switch regions

This topic describes how to switch regions managed by Apsara Stack Security.

Context

When you log on to Apsara Stack Security Center, you have selected a region. To manage the servers or network security of another region, follow these steps:

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Select the target region from the Region drop-down list in the upper-left corner.

The region of Apsara Stack Security Center is switched to the selected one.

19.4 Threat Detection Service

19.4.1 Overview

This topic describes the basic concepts of Threat Detection Service (TDS).

TDS incorporates a full range of capabilities to monitor enterprise vulnerabilities , security threats such as hacker intrusions, Web attacks, and distributed denial of service (DDoS) attacks, threat intelligence, and enterprise security reputation. Through modeling and analysis, TDS obtains key information from traffic features , server behavior, and server operations logs to identify intrusions that cannot be detected simply by inspecting traffic or scanning files. By combining the output from cloud-based analytics models with intelligence data, TDS identifies threat sources and attack behavior, and assesses the level of threat.

TDS provides the following features:

- **Overview: displays the overall security situation, network traffic, access analysis results, and security screens.**
- **Event analysis: displays security events that have been detected in the system and the event trends.**
- **Threat analysis: displays the risks that threaten the system security.**
- **Security reports: allows you to configure tasks for generating Apsara Stack security reports.**
- **Asset management: allows you to manage the server assets and NAT assets in Apsara Stack.**
- **Attack blocking settings: allows you to set features for blocking Web attacks and brute-force attacks.**

19.4.2 Security overview

19.4.2.1 View security overview information

This topic describes how to view the security trends, latest threats, and asset information of the Apsara Stack platform.

Context

The Security Overview page presents an overview of the detected security events, latest threats, and inherent vulnerabilities and defects. The security administrator

can view the information on the Security Overview page to have a comprehensive understanding of the system security situation.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. **Choose Threat Detection Service > Overview and then click the Security Overview tab.**

   View the current security situation of the Apsara Stack platform, as shown in *Figure 19-1: Security Overview tab*.

   Figure 19-1: Security Overview tab

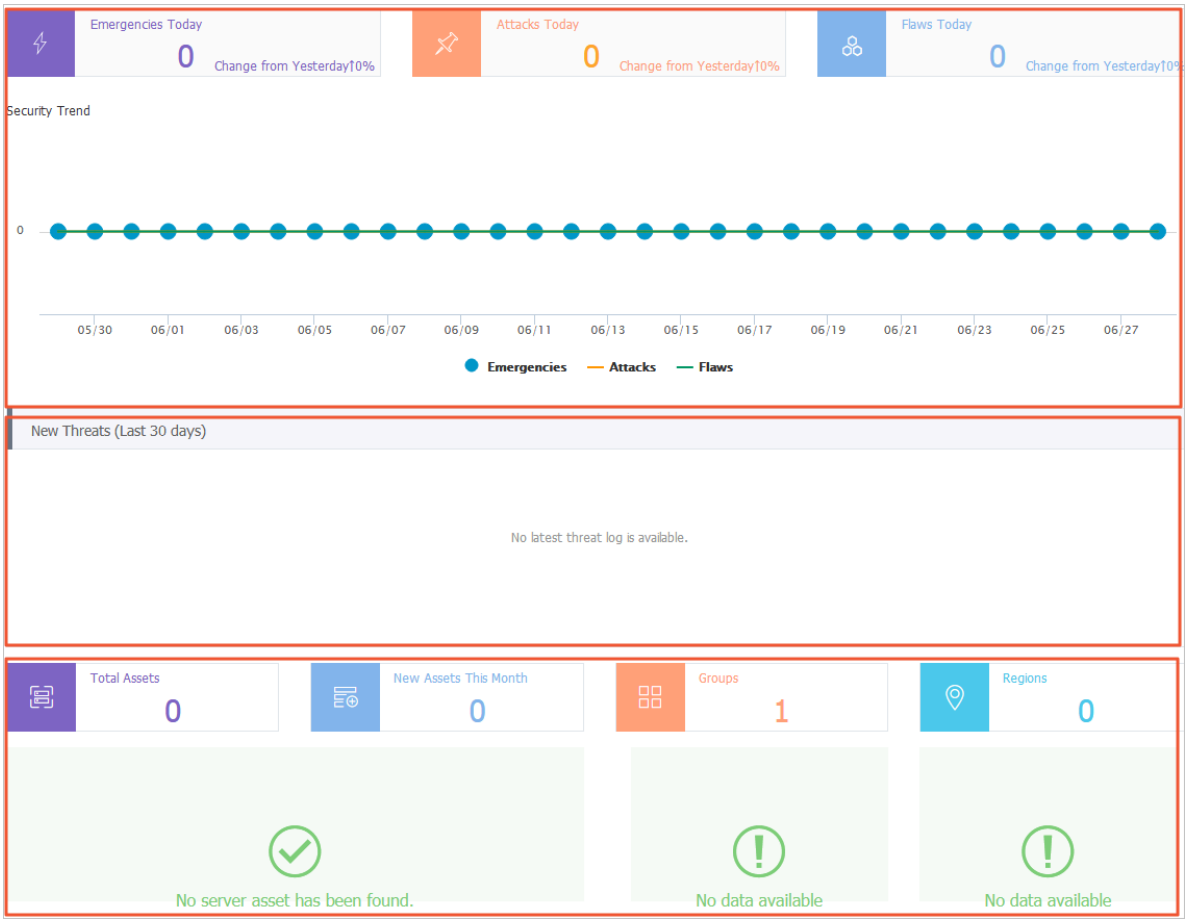


Table 19-3: Sections on the Security Overview tab

Section	Description
Security Trend	Displays the detected security events and attacks, system vulnerabilities and defects, and system security trends by time.

Section	Description
Latest Threats	Displays the existing security threats in the system. These threats require immediate attention. **Note:** The threats are identified by the core scanner of Apsara Stack Security and analyzed by the big data analysis model of Apsara Stack.
Asset Overview	Displays the information about your most important assets so that you can learn the asset status in real time.

3. Click Emergencies Today, Attacks Today, or Flaws Today to view details on the corresponding page. You can also click View for each threat in the Latest Threats section to view the details.

   For example, you can click Emergencies Today to go to the Event Analysis page.

19.4.2.2 View the network traffic information

This topic describes how to view the network traffic information.

Context

The system displays network traffic information in the specified time period in a line chart. By checking the traffic at different time points, in different regions, and from each IP address, you can identify the time when the traffic reaches the highest or lowest and view traffic distribution by rate or region. You can also check the top 5 IP addresses that generate the most traffic to effectively block access from malicious IP addresses.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose Threat Detection Service > Overview and then click the Network Traffic tab.
3. Optional: Set Region, enter the IP address of an Elastic Compute Service (ECS) instance, and then click Search.

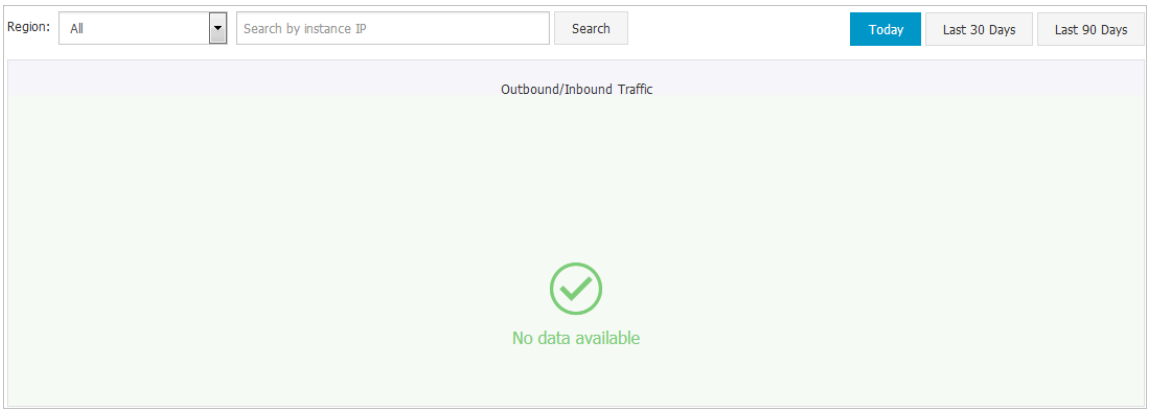
   You can query the traffic information by region and IP address.
4. Optional: Click Today, Last 30 Days, or Last 90 Days.

   You can query the traffic information in different time periods.

5. **View the traffic information at a specified time point.**

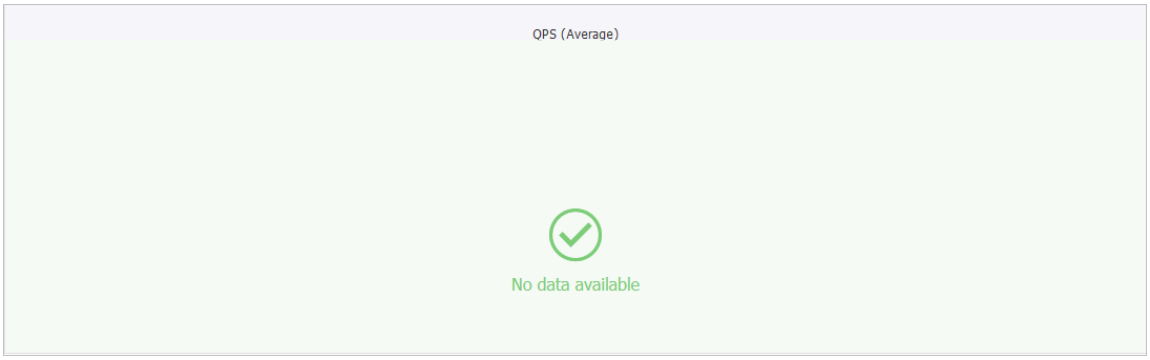
   - **In the outbound or inbound traffic diagram, move the pointer over a point on the traffic curve. You can view detailed information about the outbound or inbound traffic at the specified time point and the top 5 IP addresses that generate the most traffic.**

     Figure 19-2: View the top 5 IP addresses that generate the most traffic



   - **In the QPS (average) diagram, move the pointer over a point on the traffic curve. You can view the detailed QPS information at the specified time point.**

     Figure 19-3: View detailed QPS information



19.4.2.3 View access analysis results

This topic describes how to view access analysis results.

Context

Based on big data analysis on access from different sources, access activities are divided into the following types: normal access, malicious access, and crawler access. Based on malicious access and crawler access, the security administrator can figure out the causes of possible security issues in the system.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. **Choose Threat Detection Service > Overview and then click the Visits tab.**

   The Visits tab appears.

   Table 19-4: Sections on the Visits tab

Section	Description
All Visits Yesterday	Displays the top 10 most-accessed domain names on the previous day and the number of IP addresses that accessed each domain name.
Visitors Detected	Displays the numbers of normal, malicious, crawler access activities on the previous day.
History Details	Displays the historical information regarding malicious access and crawler access.

3. **In the History Details section, click All, Malicious IP, or Crawler IP.**

   The detailed access information appears.

   Table 19-5: Detailed access information

Parameter	Description
Visitor IP	The IP address of the visitor.
Detection Method	The access type, including malicious access and crawler access.
Time	The time of the access.
UserAgent	The User-Agent information contained in the HTTP request.
Target Application	The URL of the application that is accessed.
Visited Pages	The number of pages that have been accessed.
Maximum Visits per Second	The maximum number of access activities per second.
Web Attack Detected	Indicates whether the access involves web attacks.

19.4.2.4 View information on visualization screens

This topic describes how to view the information on visualization screens.

Context

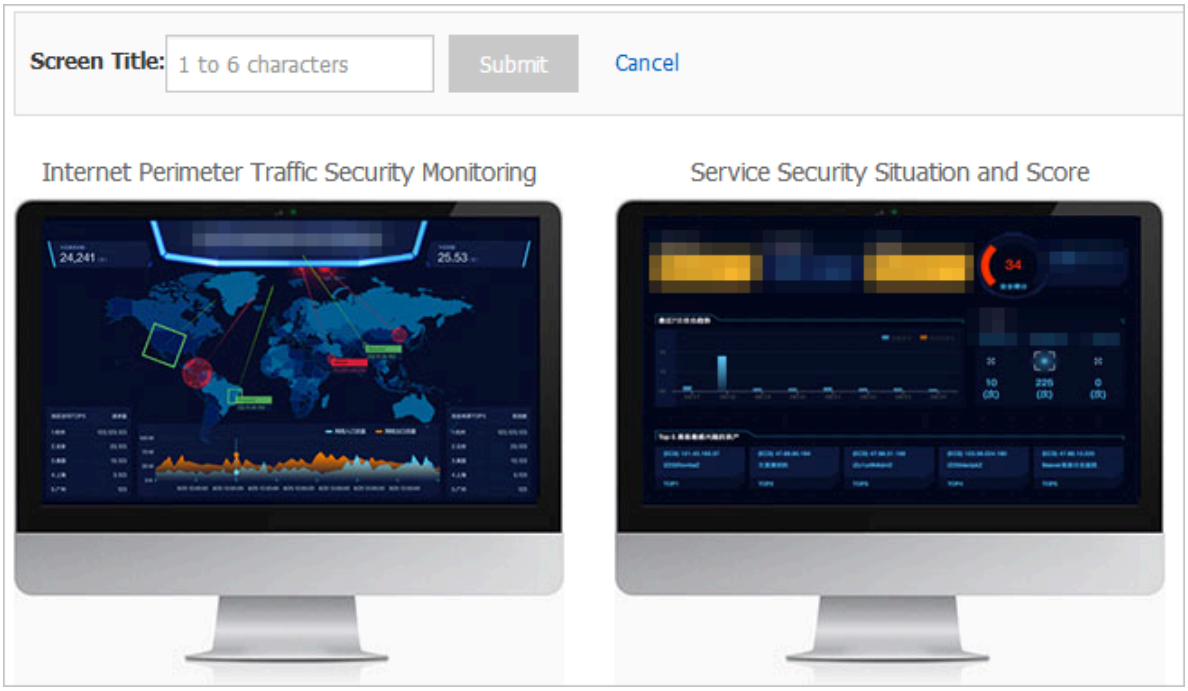
The visualization screens use animations to display key metrics of security events . This provides the security administrator with a general picture of the security situation, effectively supporting security decisions.

The visualization screens include the screen for monitoring the security of the Internet perimeter traffic and the screen for monitoring service security situation and scoring.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. **Choose Threat Detection Service > Overview and then click the Screens tab.**

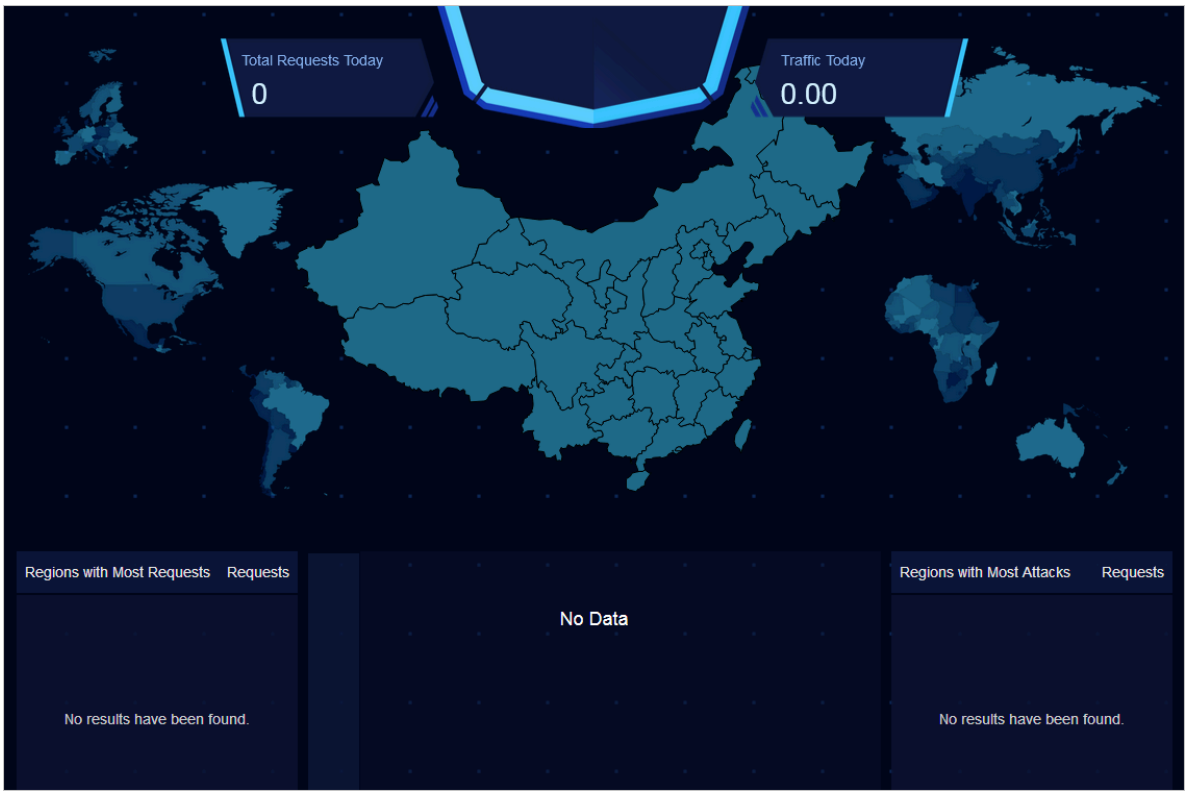
Figure 19-4: Screens tab



Click Modify on the Screens tab to modify the screen title.

3. **Click the Internet Perimeter Traffic Security Monitoring screen.**

   You can view the information on the screen.



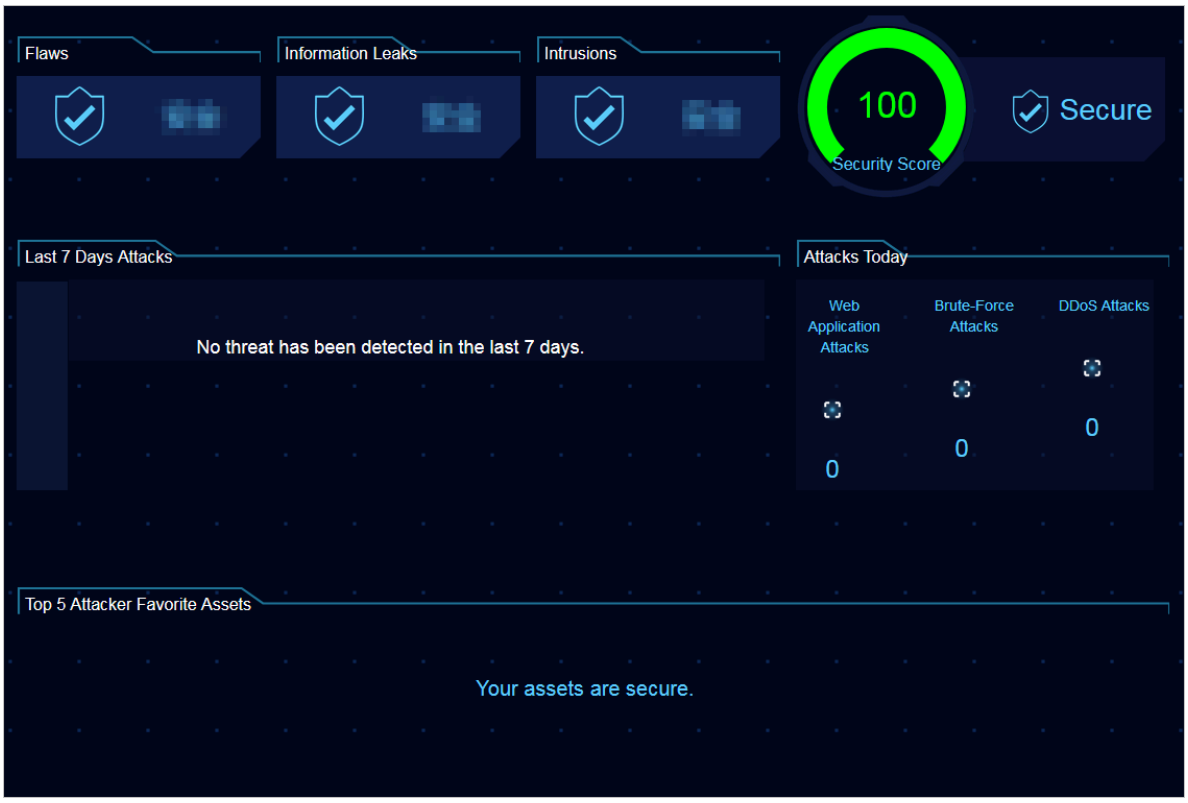
The Internet Perimeter Traffic Security Monitoring screen displays statistics on the source regions and the numbers of current requests and attacks. It also displays a general picture of the system traffic. It lists the top 5 request source regions and top 5 attacked regions, providing the security administrator with an accurate understanding of the regional distribution of requests and attacks.

Table 19-6: Access traffic data sources

Type	Implementation
Request analysis	The assets that interest users are pushed to the traffic security monitoring module, which then reports access information for these assets.
Attack analysis	The traffic security monitoring module detects, reports, and displays events suspected to be Web attacks.
Traffic display	The traffic security monitoring module collects and reports traffic information to Apsara Stack Console for recording.

4. **Click the Service Security Situation and Score screen.**

   You can view the information on the screen.



   The Service Security Situation and Score screen displays detailed information about the security events facing the system. By analyzing the system vulnerabilities or defects and the assets that have been attacked or that interest hackers, this screen evaluates the system's security situation and displays the security grade.

   The data shown on this screen is reported by modules such as traffic security monitoring, Server Guard, and defect analysis of Apsara Stack Security. The top 5 assets that interest hackers the most are analyzed by the big data engine through modeling.

19.4.3 Event analysis

19.4.3.1 View emergencies

This topic describes how to view emergencies.

Context

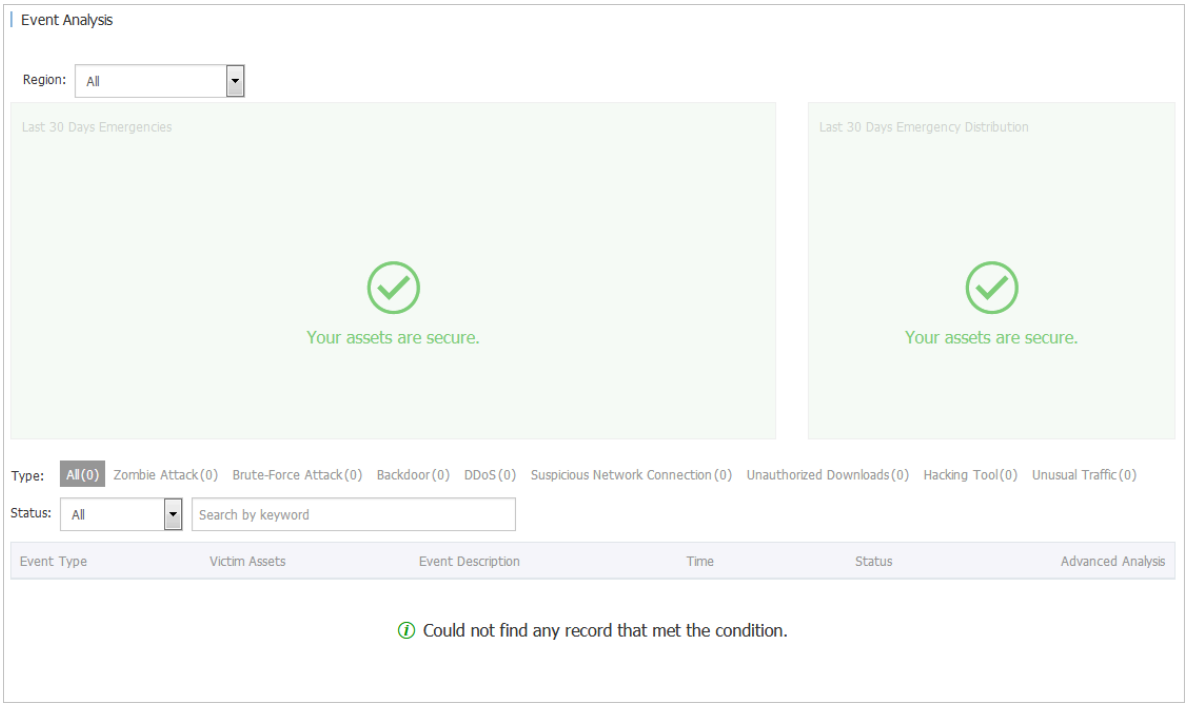
Emergencies are security events that have occurred or are currently occurring in the system. Emergencies are detected and reported by Apsara Stack Security

modules such as traffic security monitoring, Server Guard, and defect analysis. When emergencies occur, the security administrator must pay immediate attention and take appropriate security measures.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose Threat Detection Service > Event Analysis.

   Figure 19-5: Event Analysis page



3. Select a specific emergency type for Type.

   Emergencies of the specified type appear in a list.

   Table 19-7: Emergency event types

Emergency event type	Description
Zombie attack	A user server is controlled by hackers and used as a bot to launch external attacks.
Brute-force attack	The Server Guard client reports both brute-force attack attempts and successful attacks. A successful attack will appear in the emergency list. It must be immediately handled by the security administrator.

Emergency event type	Description
Webshell	The Server Guard client detects webshell files in the system . The big data analysis module analyzes traffic imported from the traffic security monitoring module to detect one-line trojans and complex trojans.
DDoS attack	Distributed denial of service (DDoS) attacks are detected by the traffic security monitoring module.
Suspicious network connections	Apsara Stack Security uses big data analysis models to analyze the information reported by security modules, and detects suspicious behavior such as suspicious external connections, malicious program downloads, and malicious file downloads.
Unauthorized download	Distinctive responses within a specified quantity range (greater than 1, less than 20) are selected from the output traffic of the traffic security monitoring module. This allows the big data analysis module to detect unauthorized downloads.
Hacking tool	The residual hacking tools or hacker attack behavior on the servers can be detected based on the information reported by Server Guard.
Unusual traffic	Attacks such as miner programs can be detected based on the information reported by the traffic security monitoring module and Server Guard.

4. View detailed event information in the list.

19.4.4 Threat analysis

19.4.4.1 View threat analysis results

This topic describes how to view threat analysis results.

Context

Apsara Stack Security analyzes traffic information by using the big data model to detect attack features, integrates attack information by attack type, and presents the current security threats in the system.

The threat analysis results cover the following information:

- **Trends of normal attacks and targeted attacks in the last 7 days and the last 30 days.**

- **Top 5 assets that interest hackers: The system analyzes traffic information by using the big data model and grades each asset by threat, and displays the five most risky assets for attention and protection by the security administrator.**
- **Targeted attack analysis: The system analyzes the traffic information provided by the traffic security monitoring module by using the big data model to detect targeted attacks.**

Procedure

1. *Log on to Apsara Stack Security Center*.
2. **Choose Threat Detection Service > Threat Analysis and then click the Threat Analysis tab.**

   The threat analysis results appear.

   Table 19-8: Threat analysis

Parameter	Description
Last 7 Days Attacks	Displays attacks to servers and applications on the Apsara Stack platform in the last 7 days.
Last 30 Days Attacks	Displays attacks to servers and applications on the Apsara Stack platform in the last 7 days.
Attacker's Top 5 Favorite Assets	Displays the top 5 assets that interest hackers. These asset IP addresses are obtained by Apsara Stack Security Center by using the big data computing model based on detected attack and threat information. We recommend that you enhance the protection on these assets.
Targeted Attacks	Displays the targeted attacks of a specified type in Apsara Stack Security Center.

3. **Select a targeted attack type for Type.**

   The targeted attack events of the specified type appear in a list. *Table 19-9: Targeted attack types* **describes targeted attack types.**

   Table 19-9: Targeted attack types

Targeted attack type	Description
Targeted Web attack	When the system discovers a targeted Web attack , this means that hackers are more interested in a website than others. The hackers have performed malicious operations such as SQL injection, command execution, or directory scan on this website.
Targeted server password cracking	This type of attack aims at cracking users' logon passwords. Hackers generally launch untargeted cracking attacks on server passwords. A targeted attack generally implies that hackers are interested in specific servers.
Credential stuffing attack	The system can analyze unusual logon activities to detect logons resembling credential stuffing attacks . Such attacks indicate that hackers may be using username and password combinations leaked on the Internet in an attempt to forcibly log on to a website . This may harm user interests.
CMS unusual logon	The system can detect unusual logon events for the application administration console. If this logon attempt is not made by the authorized user, the hacker may have already stolen the background password. In this case, we recommend that you check the password strength and change the password as soon as possible.
Scanner-based attack	The system can detect hackers' behavior of using a dedicated vulnerability scanner to scan servers on the Apsara Stack platform. After detecting a vulnerability on a server, hackers may launch a targeted attack to the server.
Pingback exploit	The system can detect targeted attacks that are launched by hackers by exploiting the vulnerability of pingback.

Targeted attack type	Description
Logon with multiple accounts	The system can detect attackers that are using a large number of low-quality accounts to log on. Such accounts are most likely bot accounts.

4. In the Targeted Attacks list, click View next to a specific attack.

   The detailed information and solution for the targeted attack appear.

19.4.4.2 View attack information

This topic describes how to view attack information.

Context

Attacks include application attacks and brute-force attacks.

- Application attacks: The traffic security monitoring module of Apsara Stack Security monitors all traffic to Web servers and extracts attack information.
- Brute-force attacks: The Server Guard client installed on a server detects hackers ' brute-force attacks to this server and reports the attacks to Apsara Stack Console.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose Threat Detection Service > Threat Analysis and then click the Attacks tab.
3. Set Region and click Application Attack.

   The application attack events appear.

   Table 19-10: Application attacks

Parameter	Description
Last 7 Days Attacks	Displays attacks to applications on the Apsara Stack platform in the last 7 days.
Attack Types of Last 7 Days	Displays the types of attacks to applications on the Apsara Stack platform in the last 7 days.

Parameter	Description
Application Attack List	Displays application attack events.

Select an application attack type in the Type section.

Attack events of the corresponding type appear in a list. *Table 19-11: Application attack types* **describes the application attack types.**

Table 19-11: Application attack types

Application attack type	Description
SQL injection	A Web application does not check the validity of the data that users provide. An attacker creates SQL statements to submit special characters and commands from an input area on the webpage, such as the address box for entering a URL or a form. In this way, the attacker interacts with the database to obtain private information or tamper with the database data.
XSS attack	A Web application does not filter or restrict the statements and variables that users provide. An attacker submits malicious code to the database or HTML page from an input area on the webpage. When users click the link or open a page that contains malicious code, malicious code automatically runs on the browser.
Code or command execution	An attacker issues requests using URLs and runs unauthorized code or commands on the Web server.
Local or remote file inclusion	An attacker adds invalid parameters to URLs when issuing requests to the Web server. The Web server fails to filter variables and uses these invalid parameters. These invalid parameters may be the names of local files or remote malicious files. This vulnerability is caused by the failure to strictly filter PHP variables. Only PHP-based Web applications may have the file inclusion vulnerability.

Application attack type	Description
Trojan script	A Trojan script is a command execution environment in the form of Web files such as ASP, PHP, and JSP. It is also known as a webshell. After intruding a website, an attacker usually mixes ASP or PHP webshell files with normal webpage files in the Web directory of the website server. Then, the attacker can access the webshell files from a browser to obtain the command execution environment for controlling the website server.
Upload vulnerability	When processing a file uploaded by a user, a Web application does not check the validity of the file name extension or the validity of the content in the file before storing the file on the server. This file may be a webshell that can control the Web server directly.
Path traversal	When issuing requests to a Web server, an attacker adds `../` and its variant to a URL or a special directory. The attacker can then access the unauthorized directory and run commands in directories except for the root directory of the Web server.
Denial of service (DoS)	An attacker uses DoS to exhaust the network or system resources on a server and interrupt or stop services on this server. This prevents authorized users from accessing the server.
Unauthorized access	An application has vulnerabilities in the authentication process. An attacker exploits these vulnerabilities to bypass authentication and access or operate unauthorized code.
Others	Other application attack types.

4. **Set Region and click Brute-Force Attack.**

   The brute-force attack events appear.

   Figure 19-6: Brute-force attack

   Select a brute-force attack event and click Show to view the details of the event.

19.4.5 Security reports

19.4.5.1 Create a report task

This topic describes how to create a report task. A report task regularly sends security reports of the Apsara Stack platform to a specified email address, informing the security administrator of the current security situation.

Context

The following table describes the information that a security report can contain.

Item	Sub-item	Description
Threat Detection Service	Security statistics	The security overview information on the overview page of Threat Detection Service (TDS).
	Highlights	The important emergency information on the event analysis page of TDS.
	Threat trend	The attack trends and analysis information on the threat analysis page of TDS.
Security protection	Distributed denial of service (DDoS)	The DDoS attack events detected by Apsara Stack Security Center.

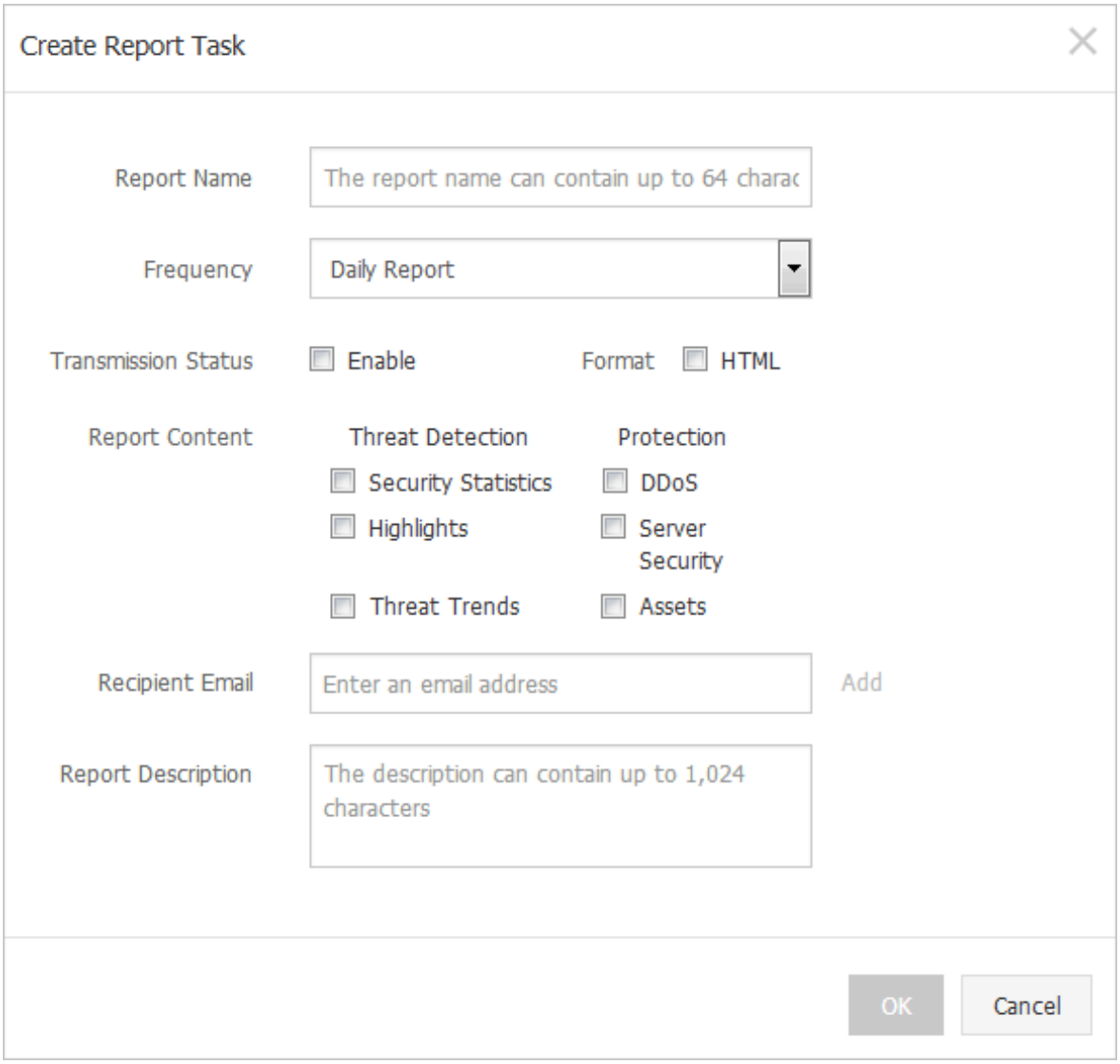
Item	Sub-item	Description
	Server security	The server security vulnerabilities, unusual logons, brute-force attacks, and configuration risks detected by Apsara Stack Security Center.
	Protected assets	The assets protected by Apsara Stack Security Center, including server assets and NAT assets.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose Threat Detection Service > Security Reports.
3. On the Security Reports page, click Create Task.

4. **In the Create Report Task dialog box, set relevant parameters.**

Figure 19-7: Create a report task



Parameter	Description
Report Name	**The name of the report task.**
Frequency	**The interval at which security reports are sent.** **Value values:** • `Daily Report`: **indicates that security reports are sent on a daily basis.** • `Weekly Report`: **indicates that security reports are sent on a weekly basis.** • `Monthly Report`: **indicates that security reports are sent on a monthly basis.**

Parameter	Description
Transmission Status	Specifies whether to enable this report task.
Format	Specifies whether to generate reports in HTML format.
Report Content	The items to be contained in a security report.
Recipient Email	The email address that receives security reports. **Note:** Click Add next to the field to add an email address. You can add up to 10 email addresses.
Report Description	The report description.

5. Click OK.

Result

After the report task is created, the specified email addresses receive security reports at the specified intervals, as shown in *Figure 19-8: Security report*.

Figure 19-8: Security report

19.4.5.2 Manage report tasks

This topic describes how to view, modify, or delete report tasks.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose Threat Detection Service > Security Reports.

3. On the Security Reports page, manage existing report tasks.
   - Select a report task and click Details to view details of this task.
   - Select a report task and click Modify to modify this task.
   - Select a report task and click Delete to delete this task.

19.4.6 Manage assets

19.4.6.1 Overview

Apsara Stack Security Center presents statistical information about your assets in charts, for example, your server assets and NAT assets, frequency of increase or decrease in the assets, and regional distribution. The security administrator can query asset information by group or type, so that they can better understand the general asset information for better asset management.

On the Asset Overview page, the security administrator can view the overall asset information in a direct and clear way, including the total number of assets, number of new assets in the current month, number of groups, number of regions, and asset distributions by report time, group, and region. This helps users better manage their assets.

Figure 19-9: Asset Overview page

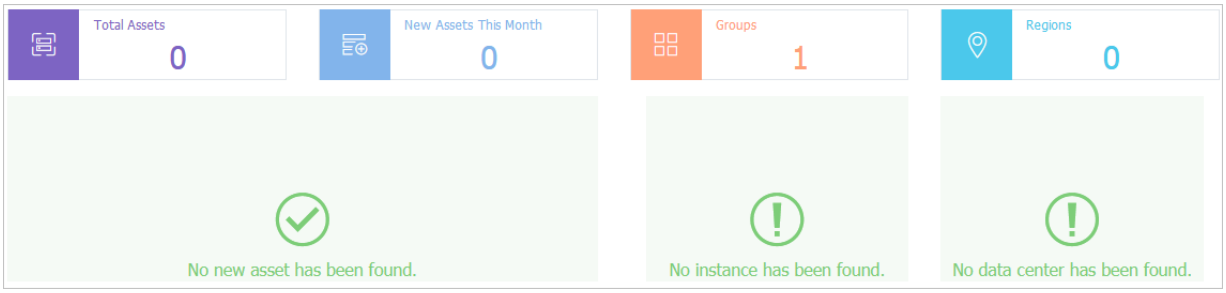


Table 19-12: Parameters on the Asset Overview page

Parameter	Description
Total Assets	The total number of assets reported by the Server Guard agent, including server assets and NAT assets.
New Assets This Month	The total number of new assets in this month, including server assets and NAT assets.
Asset Distribution by Report Time	The change in the number of server assets and that of NAT assets over the last 7 days.

Parameter	Description
Groups	The number of existing groups.
Asset Distribution by Group	The pie chart that shows the proportion of assets in each group to the total assets.
Regions	The number of configured regions.
Asset Distribution by Region	The pie chart that shows the proportion of assets in each region to the total assets.

19.4.6.2 Manage groups

19.4.6.2.1 Add a group

This topic describes how to add a group.

Context

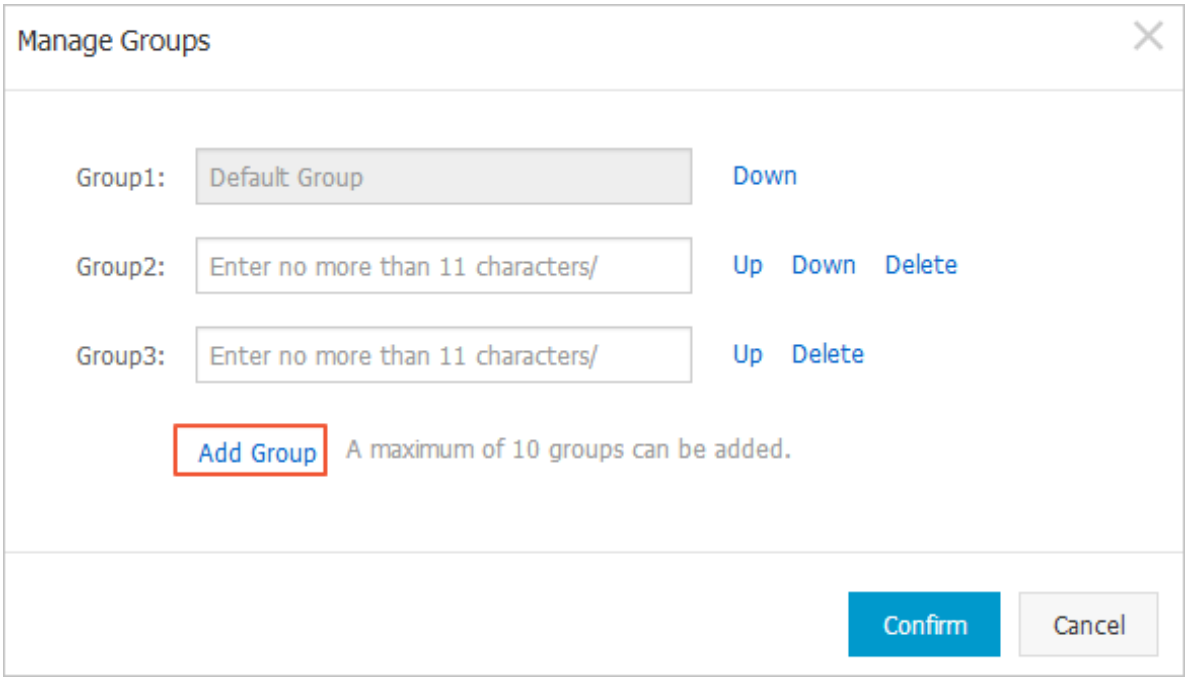
Asset groups are used to differentiate assets, making it easier for you to query and modify the asset information.

Note:
A maximum of 10 asset groups are supported.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Threat Detection > Assets.**
3. **In the Group area, click Manage Groups.**

4. In the Manage Groups dialog box, click Add Group.



5. Enter a group name, and click Confirm.

19.4.6.2.2 Delete a group

This topic describes how to delete unnecessary groups to facilitate asset information query and modification.

Context

- You cannot delete or rename the default group.
- You cannot delete groups that contain assets.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Threat Detection > Assets.
3. In the Group area, click Manage Groups.
4. In the Manage Groups dialog box, click Delete next to a group.
5. Click Confirm.

19.4.6.2.3 Sort groups

This topic describes how to sort groups. You can move frequently used groups to the top to facilitate asset information query and modification.

Procedure

1. *Log on to Apsara Stack Security Center.*

2. Choose Threat Detection > Assets.
3. In the Group area, click Manage Groups.
4. In the Manage Groups dialog box, click Up or Down to sort the groups.
5. Click Confirm to save the new group order.

19.4.6.3 Asset information

19.4.6.3.1 Manage server assets

A server asset refers to a server where a Server Guard agent has been installed and has connected to the Server Guard server.

Context

The security administrator can view the server asset information such as the operating systems, enabled ports, and installed common software. The security administrator can also change the region and group for each server asset.

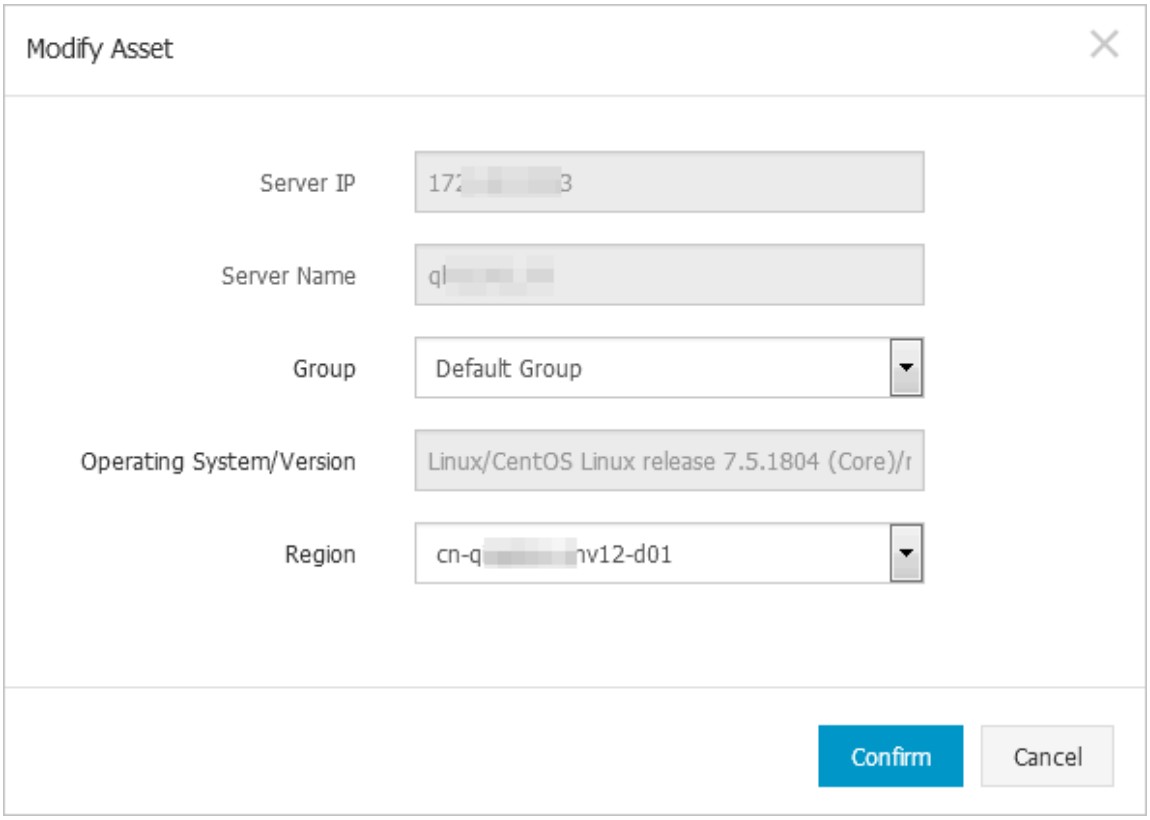
Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Threat Detection > Assets.
3. In the Server Asset/NAT Asset area, click the Server Asset tab.
4. Set the search criteria and click Search to view a server asset.

> **Note:**
> **You can filter server assets by operating system, region, or group. You can also enter a server IP address or server name for a fuzzy search. By default, the server assets in all regions are displayed and sorted by IP address.**

5. Maintain the server asset information.

   - Click Modify. In the Modify Asset dialog box, change the asset group and region, and click Confirm.



   - Click Delete. In the Delete Asset message, click Confirm to delete the asset.

> **Note:**
> If the Server Guard agent on a server is uninstalled or an ECS instance is removed from Apsara Stack, you must manually delete the corresponding asset.

19.4.6.3.2 Manage NAT assets

This topic describes how to add, view, and delete NAT assets.

Context

NAT assets are external IP addresses that are converted from internal IP addresses through NAT, namely, IP addresses that are exposed to the Internet. Multiple servers can share one external IP address but use different ports to receive Internet requests. After an IP address is set as a NAT asset, Threat Detection Service analyzes this asset to detect attack events.

The security administrator can search for a NAT asset protected by Apsara Stack Security to view the basic information about the asset or change the asset group or region. The security administrator can also add one NAT asset or add multiple NAT assets by specifying a CIDR block.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Threat Detection > Assets.
3. In the Server Asset/NAT Asset area, click the NAT Asset tab.
4. Set the search criteria and click Search to view a NAT asset.

   Note:

   You can filter NAT assets by region or group. You can also enter a NAT IP address for a fuzzy search. By default, the NAT assets in all regions are displayed and sorted by IP address.

5. Add a NAT asset.

   Note:

   The NAT IP address to be added cannot be the same as an existing IP address. Specify a valid IP address or CIDR block as the NAT IP address.

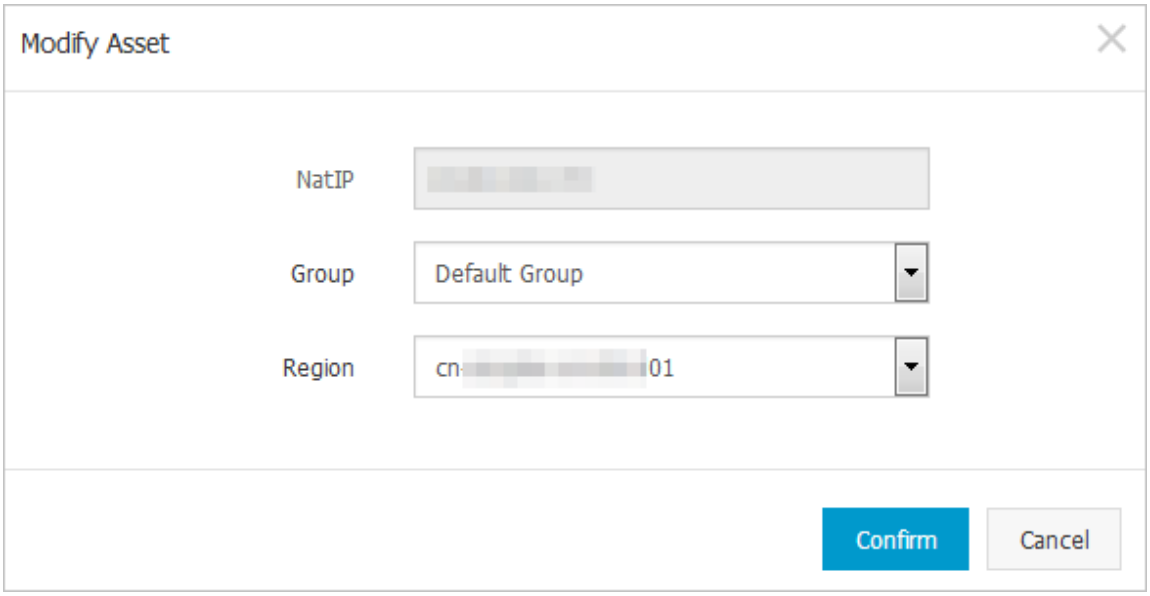
   a) Click Add in the upper-right corner of the NAT Asset tab page.

   b) In the Add Asset dialog box, enter an IP address or a CIDR block and select a group and a region.

   c) Click Confirm.

6. Maintain NAT asset information.

   - Click Modify. In the Modify Asset dialog box, change the business group and region of the asset.



   - Click Delete. In the Delete Asset message, click Confirm to delete the NAT asset.

19.4.6.3.3 Modify attributes for multiple assets

This topic describes how to modify the group and region attributes for multiple assets.

Context

You can modify the attributes for one or more assets at one time.

- **Modify an attribute for one asset.**

  This method applies when you modify only one asset or when the assets to be modified are not in the same CIDR block and use server names without similarities. For more information about how to modify one asset, see *Manage server assets* **and** *Manage NAT assets*.

- **Modify an attribute for multiple assets.**

  This method applies when you modify multiple assets that are in the same CIDR block or have similar server names.

Note:
You cannot modify the server IP address, server name, operating system, or operating system version.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Threat Detection > Assets.**
3. **Change the group or region for multiple assets.**
   - **Click Modify Group to change the group for multiple assets at one time.**
   - **Click Modify Region to change the region for multiple assets at one time.**
4. **In the Modify Group or Modify Region dialog box, specify the assets to be modified, select a new group or region for these assets, and click Confirm.**
   - **Select `CIDR Block` from the Type drop-down list, and enter the CIDR block that contains the server assets or NAT assets to be modified.**

     Note:
     If you specify a CIDR block, all server assets and NAT assets in the specified CIDR block are modified.
   - **Select `Server Name` from the Type drop-down list, and enter the common part of the names of servers to be modified.**

     Note:
     If you specify the common part of server names, all servers whose names contain the specified common part are modified.

19.4.7 Enable attack blocking

This topic describes how to enable attack blocking.

Context

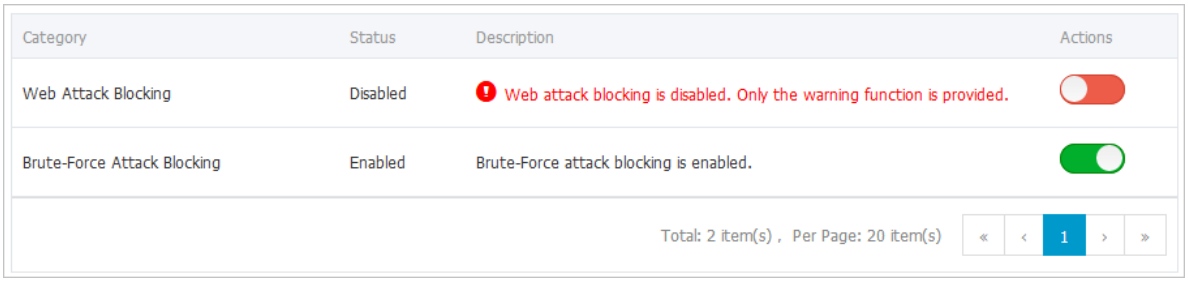
The attack blocking functions protect your servers against Web attacks and brute-force attacks.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Threat Detection > Protection Settings.**

3. **Click the toggle in the Actions column for Web Attack Blocking or Brute-Force Attack Blocking to enable or disable the corresponding function.**

   Figure 19-10: Configure attack blocking



Note:

In the Actions column, a red toggle indicates a disabled function, and a green toggle indicates an enabled function.

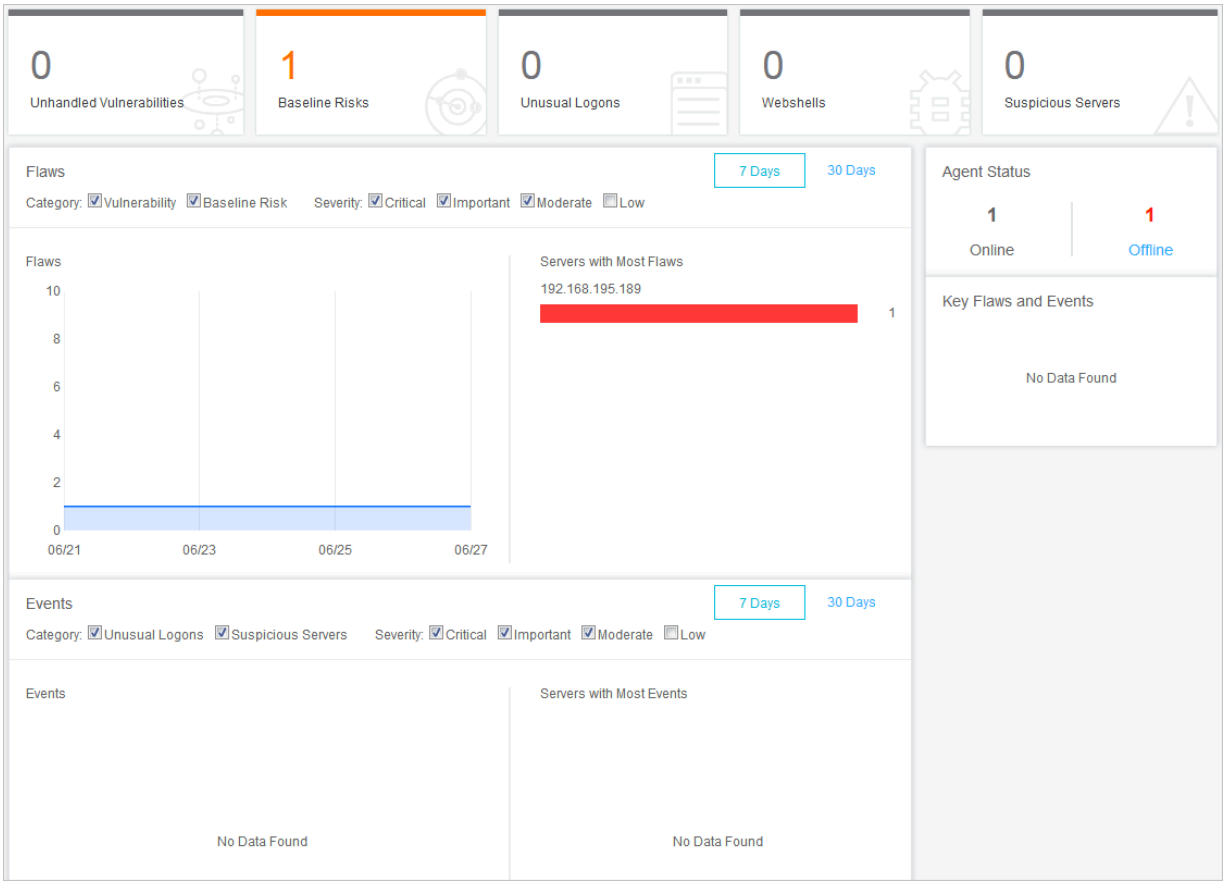
After you disable the blocking function for an attack type, Security Center provides the alert function on the attacks only.

19.5 Server security

19.5.1 Server security overview

The security administrator can view the current security status of all servers on the server security overview page of Apsara Stack Security Center.

The server security overview page contains the following areas: Overview, Flaws, Events, Agent Status, and Key Flaws and Events.



- **Overview: This area displays the number of security flaws of each type, such as unhandled vulnerabilities and baseline risks, and the number of security events of each type, such as unusual logons, webshells, and suspicious servers.**
- **Flaws: This area displays the trend of security flaws on your servers. Your server security is threatened if you do not handle the security flaws.**
- **Events: This area displays the trend of security events on your servers. A security event is an intrusion event that has been detected on your server.**
- **Agent Status: This area displays the number of servers being protected and the number of servers with an offline agent.**
- **Key Flaws and Events: This area displays the recent key flaws and events on your servers. You can click a flaw or event to view the details.**

19.5.2 Server list

19.5.2.1 Manage the server list

On the Servers page, you can view the status of servers protected by Server Guard.

Context

The protection status of a server can be:

- Online: Server Guard provides security protection for this server.
- Offline: Server Guard cannot provide security protection for this server because the Server Guard agent on this server has gone offline.
- Disable Protection: Security protection is temporarily disabled for this server. For more information, see *Disable protection*.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Server Security > Servers.
3. Optional: Search for a server.

   To view the agent status of a server, enter the server IP address in the search box, and click Search. The detailed server information, including security information, is displayed.
4. View the agent status and detailed security information of the server.

   Click  in the upper-right corner of the page to select the information columns to be displayed. The following table lists the information categories.

Category	Information
Basic information	· Server IP/name · Tag · Operating system · Region
Agent status	Agent status
Threat prevention	· Vulnerabilities · Baseline risks
Intrusion detection	· Unusual logons · Webshells · Suspicious servers
Server fingerprints	· Number of processes · Number of ports · Root account or all accounts

5. Manage servers.

Function	Actions
Change Group	Select servers and click Change Group to add the selected servers to a new group. For more information, see *Manage server groups*.
Modify Tag	Select servers and click Modify Tag to modify tags for the servers.
Security Inspection	Select servers and click Security Inspection to select the items to be checked.
Delete External Servers	Select external servers, and choose More > Delete External Servers.
Disable Protection	Select servers of which the agent status is Online, and choose More > Disable Protection. This temporarily disables protection for the selected servers to reduce server resource consumption.
Enable Protection	Select servers of which the agent status is Disable Protection, and choose More > Enable Protection. This enables protection for the selected servers.

19.5.2.2 Manage server groups

To facilitate the security control of specific servers, you can group the servers and view security events by server group.

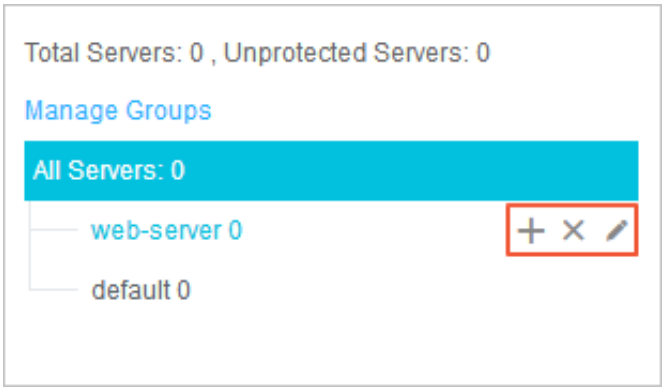
Context

Servers that have not been added to any group are assigned to the default group. If you delete a group, all servers in this group are moved to the default group automatically.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Server Security > Servers.

3. Manage server groups.



- Create a group.

  Click the Add Subgroup icon next to All Servers or a specific group, enter the name of the group to be added, and click OK.

  > **Note:**
  > The system supports three levels of groups.

- Modify a group.

  Click the Modify Group Name icon next to the target group, enter a new name, and click OK.

- Delete a group.

  Click the Delete icon next to the target group. In the message that appears, click OK.

  > **Note:**
  > After you delete a group, all servers in this group are moved to the default group automatically.

4. Group servers.
   a) Select servers from the list on the right.
   b) Click Change Group.
   c) In the Change Group dialog box, select a group from the drop-down list.
   d) Click OK.
5. Sort groups.

   Click Manage Groups to sort groups in descending order by priority.

19.5.3 Threat protection

19.5.3.1 Vulnerability management

19.5.3.1.1 Manage Linux software vulnerabilities

This topic describes how to manage Linux software vulnerabilities.

Context

Apsara Stack Security automatically scans the software that has been installed on your servers for vulnerabilities on the Common Vulnerabilities and Exposures (CVE) list and sends you alerts about the detected vulnerabilities. Apsara Stack Security also provides commands to fix vulnerabilities and allows you to verify these vulnerability fixes.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Server Security > Threat Prevention > Vulnerabilities, and click the Linux Software Vulnerabilities tab.**
3. **View the detected Linux vulnerabilities.**

   Note:
   You can quickly locate a vulnerability by using the search and filter functions.
4. **Click a vulnerability to go to the vulnerability details page. You can view detailed vulnerability information and affected assets on this page.**

   Note:
   You can quickly locate specific affected assets by using the search and filter functions.

   - **Basic Information: The basic information of the vulnerability, including the name, CVSS score, description, and resolution.**
   - **Affected Assets: The servers that are affected by the vulnerability.**

5. **Select an action based on the impact of the vulnerability.**

   Table 19-13: Actions on vulnerabilities

Action	Description
Generate Fix Command	Select this option to generate the commands for fixing the vulnerability. You can then log on to the server to run these commands.
Fix Now	Select this option to fix the vulnerability directly.
Restarted and Verified	If a vulnerability fix takes effect only after a server reboot, you should reboot the server only after the status of the vulnerability changes to Fixed (To Be Restarted). After the reboot, click Restarted and Verified.
Ignore	Select this option to ignore the vulnerability. The system does not alert you about ignored vulnerabilities.
Verify	Click Verify to verify the vulnerability fix. If you do not manually verify the fix, the system automatically verifies the fix within 48 hours.

   You can manage a vulnerability on one or more affected assets at one time.

   - **To manage a vulnerability on one asset, select an action from the Actions column of this asset.**
   - **To manage a vulnerability on one or more assets, select one or more affected servers, and select an action in the lower-left corner.**

19.5.3.1.2 Manage Windows vulnerabilities

This topic describes how to manage Window vulnerabilities.

Context

Apsara Stack Security automatically checks if your servers have the latest Microsoft updates installed, and notifies you of the detected vulnerabilities. Apsara Stack Security also automatically detects and fixes major vulnerabilities on your servers.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Server Security > Threat Prevention > Vulnerabilities, and click the Windows Vulnerabilities tab.**

3. Check the detected Windows vulnerabilities.

   Note:
   You can quickly locate a vulnerability by using the search and filter functions.

4. Click a vulnerability to go to the vulnerability details page. You can view detailed vulnerability information and affected assets on this page.

   Note:
   You can quickly locate specific affected assets by using the search and filter functions.

   - Basic Information: The basic information of the vulnerability, including the name, CVSS score, description, and resolution.
   - Affected Assets: The servers that are affected by the vulnerability.

5. Select an action based on the impact of the vulnerability. *Table 19-14: Actions on vulnerabilities* describes the actions.

   Table 19-14: Actions on vulnerabilities

Action	Description
Fix Now	Select this option to fix the vulnerability directly. The system caches an official Windows patch in the cloud for your server to download and update.
Ignore	Select this option to ignore the vulnerability. The system does not alert you about ignored vulnerabilities.
Verify	Click Verify to verify the vulnerability fix.
Restarted and Verified	If a vulnerability fix takes effect only after a server reboot, you should reboot the server only after the status of the vulnerability changes to Fixed (To Be Restarted). After the reboot, click Restarted and Verified.

   You can manage a vulnerability on one or more affected assets at one time.

   - To manage a vulnerability on one asset, select an action from the Actions column of this asset.
   - To manage a vulnerability on one or more assets, select one or more affected servers, and select an action in the lower-left corner.

19.5.3.1.3 Manage Web CMS vulnerabilities

This topic describes how to manage Web CMS vulnerabilities.

Context

The Web CMS vulnerability detection feature obtains the information of the latest vulnerabilities and provides patches in the cloud. This helps you quickly detect and fix vulnerabilities.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Server Security > Threat Prevention > Vulnerabilities, and click the Web CMS Vulnerabilities tab.**
3. **View all vulnerabilities.**

   Note:
   You can quickly locate a vulnerability by using the search and filter functions.
4. **Click a vulnerability to go to the vulnerability details page. You can view detailed vulnerability information and affected assets on this page.**

   Note:
   You can quickly locate specific affected assets by using the search and filter functions.
5. **Select an action based on the impact of the vulnerability.** *Table 19-15: Actions on vulnerabilities* **describes the actions.**

   Table 19-15: Actions on vulnerabilities

Action	Description
Fix Now	Select this option to fix the Web CMS vulnerability by replacing the Web files that contain the vulnerability on your server. **Note:** Before fixing the vulnerability, we recommend that you back up the Web files affected by this vulnerability. For more information about the paths of the Web files, see the paths specified in the vulnerability remarks.

Action	Description
Ignore	Select this option to ignore the vulnerability. The system does not alert you about ignored vulnerabilities.
Verify	Click Verify to verify the vulnerability fix. If you do not manually verify the fix, the system automatically verifies the fix within 48 hours.
Undo Fix	For vulnerabilities that have been fixed, click Undo Fix to restore the Web files that have been replaced.

You can manage a vulnerability on one or more affected assets at one time.

- To manage a vulnerability on one asset, select an action from the Actions column of this asset.
- To manage a vulnerability on one or more assets, select one or more affected servers, and select an action in the lower-left corner.

19.5.3.1.4 Manage other vulnerabilities

This topic describes how to manage other vulnerabilities.

Context

Apsara Stack Security automatically detects vulnerabilities on servers, such as the Redis unauthorized access vulnerability and Struts S2-052 vulnerability, and sends vulnerability alerts. After you fix a vulnerability, you can also verify whether the fix is successful.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Server Security > Intrusion Prevention > Vulnerabilities, and click the Others tab.
3. View all vulnerabilities.

   You can quickly locate a vulnerability by using the search and filter functions.
4. Click a vulnerability to go to the vulnerability details page. You can view detailed vulnerability information and affected assets on this page.

   You can quickly locate specific affected assets by using the search and filter functions.

5. **Select an action based on the impact of the vulnerability.** *Table 19-16: Actions on vulnerabilities* **describes the actions.**

   Follow the instructions to manually fix the vulnerabilities on the Others tab page.

   Table 19-16: Actions on vulnerabilities

Action	Description
Ignore	Select this option to ignore a vulnerability. The system does not alert you about an ignored vulnerability.
Verify	Click Verify to verify the fix after you have manually fixed a vulnerability. If you do not manually verify a fix, the system automatically verifies the fix within 48 hours after the vulnerability fix is complete.

   You can fix a vulnerability for one or multiple affected assets at one time.

   - **To fix a vulnerability for one affected asset, select an action from the Actions column of the asset.**
   - **To fix a vulnerability for one or multiple affected assets, select the target servers, and select an action in the lower-left corner.**

19.5.3.1.5 Configure vulnerability management

You can enable or disable automatic detection for different types of vulnerabilities, and enable vulnerability detection for specific servers. You can also set a time duration for which invalid vulnerabilities are retained, and configure a vulnerability whitelist.

Context

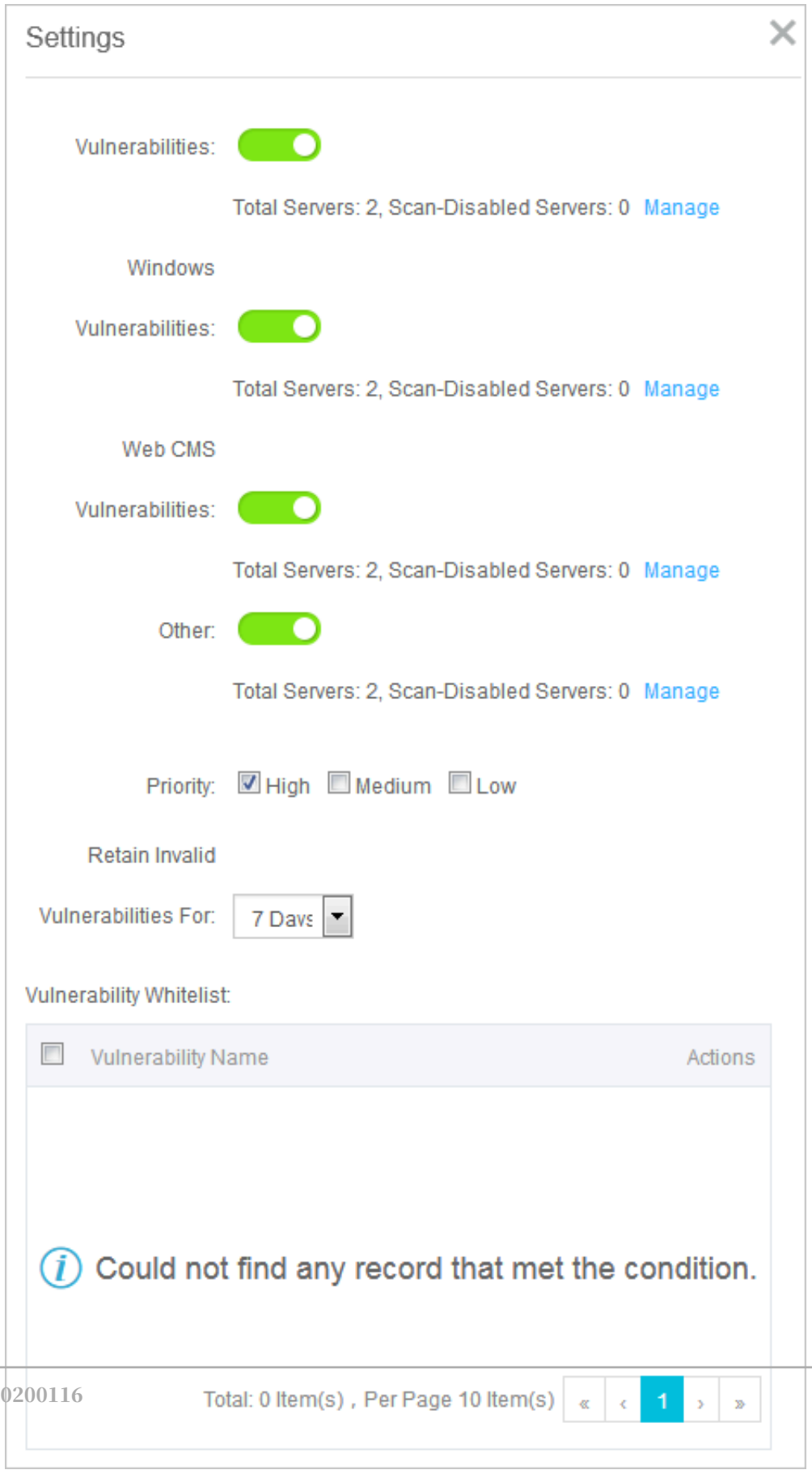
A vulnerability whitelist allows you to exclude vulnerabilities from the detection list. You can add multiple vulnerabilities in the vulnerability list to the whitelist . The system does not detect whitelisted vulnerabilities. You can manage the vulnerability whitelist on the vulnerability management settings page.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Server Security > Intrusion Prevention > Vulnerabilities.**

3. **Click Settings in the upper-right corner to configure vulnerability management policies.**

Figure 19-11: Settings dialog box



- Select a vulnerability type, and enable or disable detection for vulnerabilities of this type.
- Click Manage next to a vulnerability type and specify the servers on which vulnerabilities of this type are detected.
- Select the priorities of vulnerabilities to be detected. The priorities include `high`, `medium`, and `low`.
- Select a time duration for which invalid vulnerabilities are retained: 7 days, 30 days, or 90 days.

  Note:
  If you do not take any action on a detected vulnerability, the system determines that the alert is invalid. The system deletes the vulnerability after the specified duration.
- Select vulnerabilities in the whitelist, and click Remove to enable the system to detect these vulnerabilities and send alerts again.

19.5.3.2 Baseline check

19.5.3.2.1 Overview

The baseline check feature automatically checks the security configurations on servers, and provides the detailed check results and suggestions for baseline reinforcement.

Features

After you enable the baseline check feature, Apsara Stack Security automatically checks for risks related to the operating systems, accounts, databases, passwords, and security compliance configurations of your servers, and provides reinforcement suggestions. For more information, see *Baseline check items*.

By default, a full baseline check is performed automatically from 0:00 to 6:00 every day. You can create and manage the scan policies. When you create or modify a policy, you can customize the check items, interval, and time period of a baseline check, and select the servers to which you want to apply this policy. For more information, see *Add a custom baseline check policy*.

Notes

The following check items are disabled by default. To check these items, make sure that these items do not affect your business and select them when you customize a scan policy.

- **Weak password check for specific applications such as MySQL, PostgreSQL, and SQL Server**

  Note:
  When you run these check items, you attempt to log on to servers with weak passwords. The logon attempts consume server resources and generate many logon failure records.

- **Check items related to classified protection**
- **Check items related to the Center for the Internet Security (CIS) standard**

Baseline check items

Category	Check item
Database	Memcached security baseline check
	Redis security baseline check
Operating system	Security baseline check based on the Alibaba Cloud standard · Windows Server 2008 R2 · Windows Server 2012 R2 · Windows Server 2016 R2 · Ubuntu · Debian Linux 8 · CentOS Linux 6 · CentOS Linux 7
	Security baseline check based on the CIS standard · Windows Server 2008 R2 · Windows Server 2012 R2 · Windows Server 2016 R2 · Ubuntu 14 · Ubuntu 16 · Debian Linux 8 · CentOS Linux 6 · CentOS Linux 7

Category	Check item
	Compliance baseline check based on Grade II Protection of Information Security · Windows Server 2008 R2 · Windows Server 2012 R2 · Windows Server 2016 R2 · Ubuntu · Debian Linux 8 · CentOS Linux 6 · CentOS Linux 7
	Compliance baseline check based on Grade III Protection of Information Security · Windows Server 2008 R2 · Windows Server 2012 R2 · Windows Server 2016 R2 · Ubuntu · Debian Linux 8 · CentOS Linux 6 · CentOS Linux 7
Weak password	Weak password check for Linux
	Anonymous FTP logon check
	Weak password check for Microsoft SQL Server
	Weak password check for MySQL
	Weak password check for PostgreSQL
	Weak password check for Windows
	Weak password check for FTP
Middleware	Apache Tomcat security baseline check

19.5.3.2.2 Add a custom baseline check policy

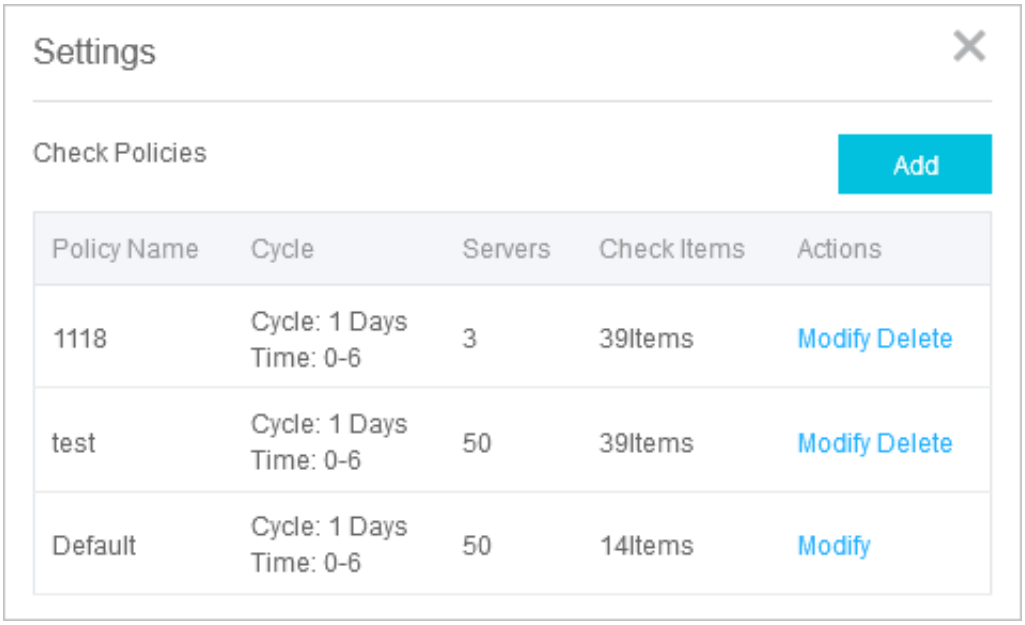
This topic describes how to add a custom baseline check policy.

Context

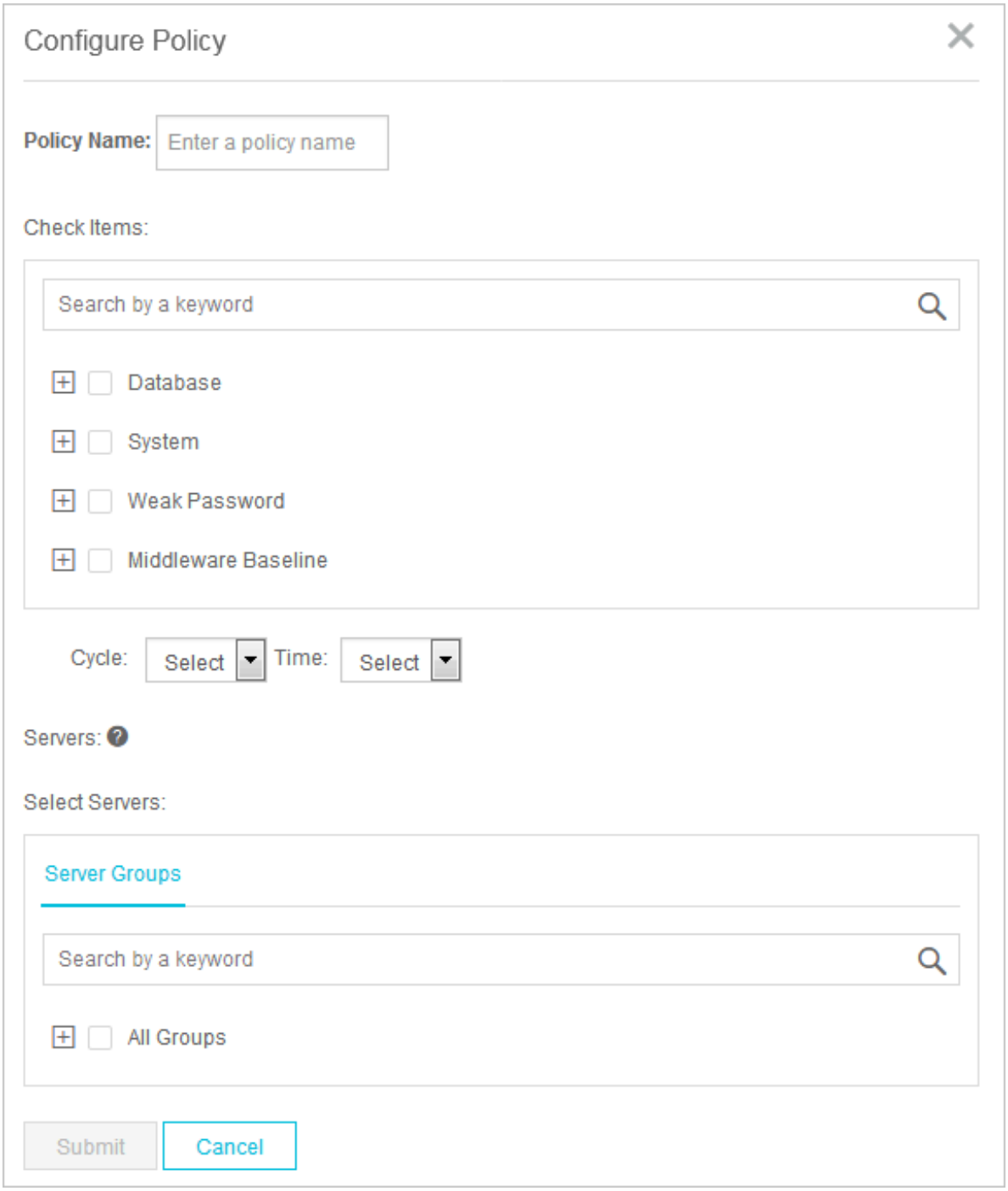
By default, the baseline check feature uses the default policy to check the baseline security of assets. You can also customize baseline check policies based on your business needs, for example, to check the compliance with Grade II Protection of Information Security.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. **Choose Server Security > Threat Prevention > Baseline Check.**
3. **On the Baseline Check page, click Settings in the upper-right corner.**
4. **In the Settings window, click Add.**



5. **In the Configure Policy window, set the policy parameters.**



Parameter	Description
Policy Name	The name of the custom policy.
Check Items	The items that the custom policy checks. For more information, see *Baseline check items*.

Parameter	Description
Cycle	The check interval and period. · Check interval: the frequency that a baseline check is performed. Valid values: `1 Day`, `3 Days`, `7 Days`, and `30 Days`. · Check period: the period during which a baseline check is performed. Valid values: `00:00-06:00`, `06:00-12:00`, `12:00-18:00`, and `18:00-24:00`.
Servers	The group of servers on which the baseline check is performed. · By default, newly created servers are in the Default group. To apply the policy to newly created servers, you must select the Default group. · For more information about how to manage server groups, see *Manage server groups*.

6. Click Submit.

19.5.3.2.3 Manage baseline check settings

This topic describes how to manage baseline check settings.

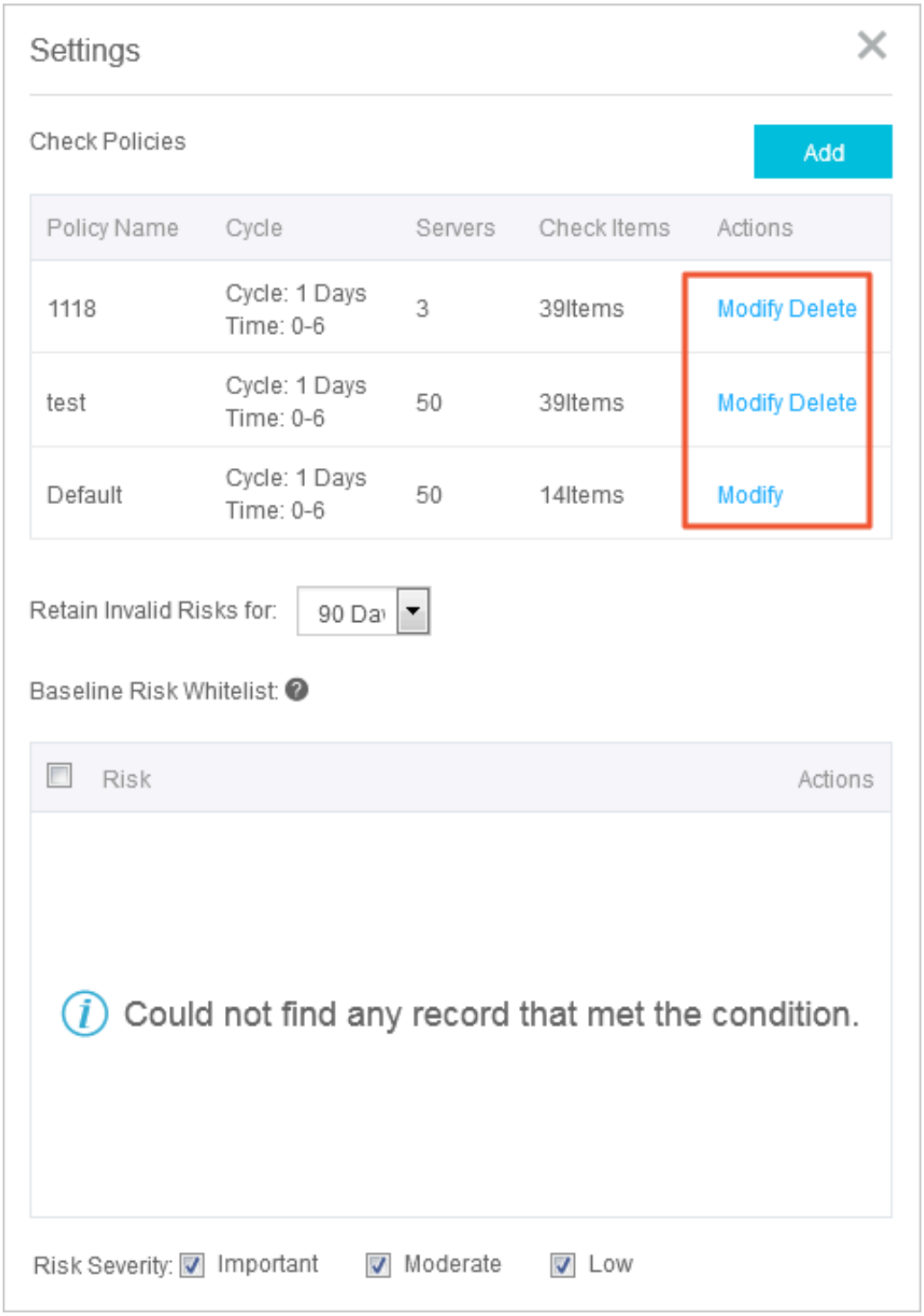
Context

You can manage baseline check settings, such as the scan policies, baseline whitelist, and baseline risk levels.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose Server Security > Threat Prevention > Baseline Check.
3. On the Baseline Check page, click Settings in the upper-right corner.

4. In the Settings window, manage the scan policies.



- Delete a scan policy.

  In the Actions column of the target policy, click Delete.

- Edit a scan policy.

  In the Actions column of the target policy, click Edit. For more information about how to set the parameters, see *Add a custom baseline check policy*.

Note:
You cannot delete the default policy or modify the check items of the default policy. However, you can modify the server group to which the default policy applies.

5. **Set Retain Invalid Risks for.**

   Select a time period from the drop-down list. Valid values: `90 Days`, `30 Days`, **and** `7 Days`.

6. **Manage the check items in the Baseline Risk Whitelist section.**

   The baseline check feature does not check the items added to the whitelist. To check an item that has been added to the whitelist, click Remove in the Actions column of the item to remove it from the whitelist.

7. **Set Risk Severity.**

   You can set Risk Severity to `Important`, `Moderate`, **or** `Low`. **After you set Risk Severity, the baseline check feature only reports the baseline risks of the corresponding risk level.**

   For example, if you set Risk Severity to `Important`, **only the baseline risks of the Important risk level appear on the Baseline Check tab.**

19.5.3.2.4 View baseline check results and resolve baseline risks

This topic describes how to view baseline check results and resolve baseline risks.

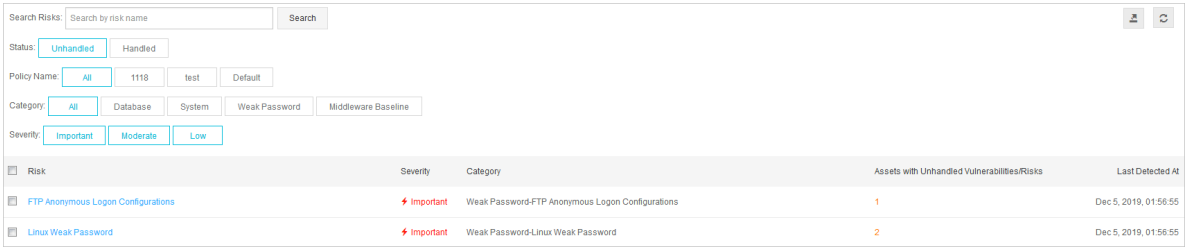
Context

A security baseline is a series of security configuration standards used to identify the basic protection capabilities of devices and systems in a network environment.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. **Choose Server Security > Threat Prevention > Baseline Check.**

3. View baseline check results.



- Search for risks.

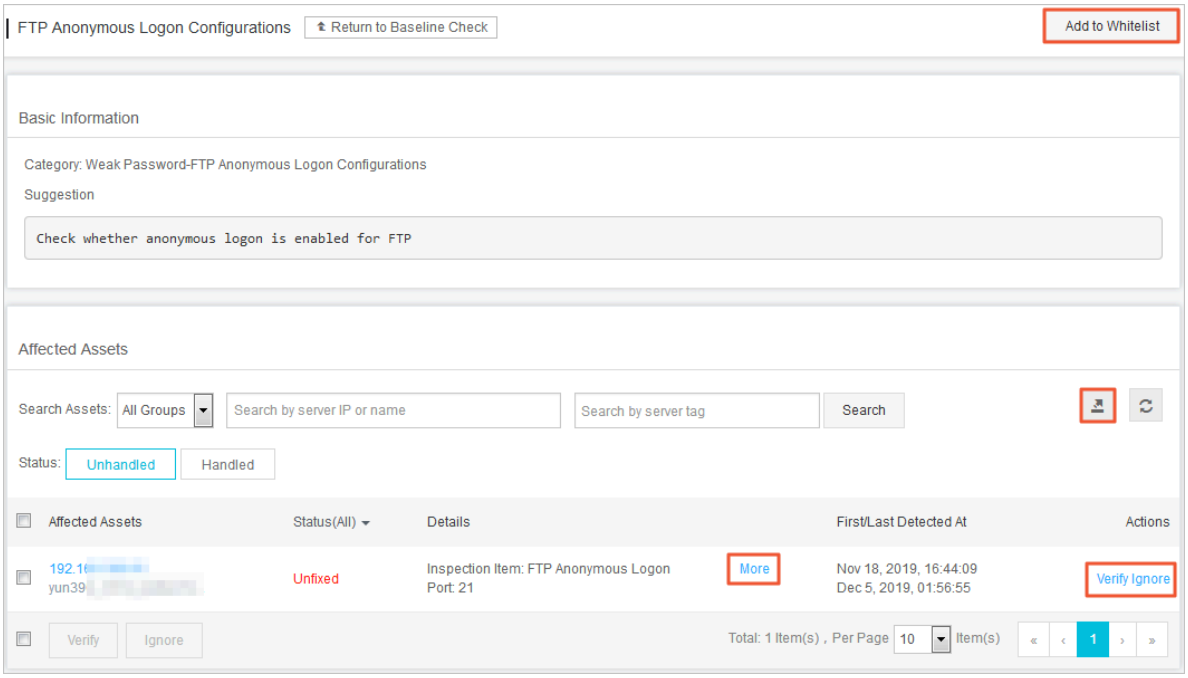
  Enter a keyword in the Search Risks field to search for risks.

- Filter risks.

  Set Status, Policy Name, Category, and Severity to filter risks.

> **Note:**
> For more information about how to set Severity, see the method for setting the Risk Severity parameter in *Manage baseline check settings*.

4. Manage a single baseline risk.



Determine the impact of the risk on your servers.

- To export the list of affected assets, click .
- If you are certain that this risk does not affect the security of your servers and you do not want the system to report the risk in the future, click Add to Whitelist in the upper-right corner.

> **Note:**
> If you want the system to report the risk again, you can remove the risk from the whitelist. For more information, see the method for setting the Baseline Risk Whitelist parameter in *Manage baseline check settings*.

- If the risk does not affect the security of some servers, click Ignore.
- If the risk affects the security of your servers, follow these steps to reinforce the servers:

a) Click the name of the baseline risk.

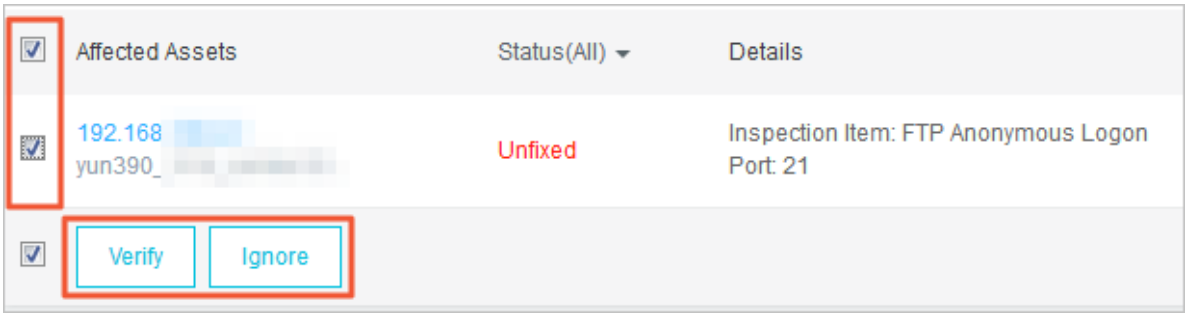
b) View the basic information about the baseline risk and the affected assets.

c) In the Details column, click More to view the check items and reinforcement suggestions.

d) Manually reinforce the relevant servers based on the reinforcement suggestions.

e) After reinforcing the servers, click Verify to check the reinforcement result.

5. Verify or ignore a risk on multiple servers at the same time.



- If a risk does not affect the security of some servers, select these servers and click Ignore to ignore the risk on these servers.
- If the affected servers have been reinforced, select these servers and click Verify to check the security reinforcement results on these servers.

19.5.4 Intrusion detection

19.5.4.1 Unusual logons

19.5.4.1.1 How unusual logon detection works

On the Unusual Logons page of the Server Guard console, you can view the IP address, account name, and time of each unusual logon. You can also view the alerts for unusual logons, disapproved IP addresses, disapproved logon time, and disapproved accounts.

The Server Guard agent regularly collects logon logs of your server, and uploads them to the Server Guard server where the logs are analyzed and matched. An alert is reported when Server Guard detects a successful logon from a disapproved location, using a disapproved IP address or account, or at a disapproved time.

Note:
To enable SMS notification, choose System Settings > Alert Settings, and then choose Logon Security > Unusual Logons to set your preferred notification methods. Value options include mobile number and email. By default, both methods are selected.

You can also set approved logon IP addresses, logon time period, and accounts for specific servers. All logon attempts, except for those using the approved logon IP addresses and accounts during the approved logon time period, will trigger alerts

. These logon security settings have a higher priority than the unusual logon alert policy.

19.5.4.1.2 Check unusual logon alerts

This topic describes how to check the alerts for unusual logons, including logons from disapproved locations, brute-force attacks, logons using disapproved IP addresses, logons using disapproved accounts, and logons at a disapproved time.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Server Security > Intrusion Detection > Unusual Logons.**
3. **Check the unusual logon alerts.**

   You can quickly locate a specific unusual logon alert by using the search and filter functions.
4. **Handle unusual logon alerts.**

   Select an unusual logon alert to check whether it is a false positive.

   - **If this alert is a false positive, click Label as Handled.**
   - **If the logon is an intrusion, improve the security of the related server. For example, use a more complex password, fix vulnerabilities on the server, remove baseline risks, or specify a blacklist or a whitelist. Then, click Label as Handled.**

19.5.4.1.3 Configure logon security

This topic describes how to configure logon security. You can set approved logon IP addresses, approved logon time periods, and approved accounts.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Server Security > Intrusion Detection > Unusual Logons.**
3. **In the upper-right corner of the Unusual Logons tab page, click Logon Security.**

4. Set approved logon locations.

   You can take the following steps to add approved logon locations:

   a) Click Add.

   b) Select a logon location from the drop-down list.

   c) Specify the servers to which the selected approved logon location applies.

      - Click All Servers to select specific servers.
      - Click Server Groups to select servers by group.

   d) Click OK.

   Click Modify or Delete to modify or delete an approved logon location.

5. Set approved logon IP addresses.

   You can enable or disable the Disapproved IP Alert toggle. The toggle is enabled if it turns green.

   Take the following steps to add an approved logon IP address.

   a) Click Add.

   b) In the Specify an Approved Logon IP area, enter a logon IP address.

   c) Specify the servers to which the specified approved IP address applies.

      - Click All Servers to select specific servers.
      - Click Server Groups to select servers by group.

   d) Click OK.

   Click Modify or Delete to modify or delete an approved logon IP address.

6. Set approved logon time periods.

   You can enable or disable the Disapproved Time Alert toggle. The toggle is enabled if it turns green.

   Take the following steps to add approved logon time periods:

   a) Click Add.

   b) In the Specify an Approved Logon Duration area, specify a time period.

   c) Specify the servers to which the specified approved logon time period applies.

      - Click All Servers to select specific servers.
      - Click Server Groups to select servers by group.

   d) Click OK.

   Click Modify or Delete to modify or delete an approved logon time period.

7. Set approved accounts.

   You can enable or disable the Disapproved Account Alert toggle. The toggle is enabled if it turns green.

   Take the following steps to add approved accounts:

   a) Click Add.

   b) In the Specify an Approved Account area, enter an account.

   c) Select servers to which the specified approved account applies.

   - Click All Servers to select specific servers.
   - Click Server Groups to select servers by group.

   d) Click OK.

   Click Modify or Delete to modify or delete an approved account.

19.5.4.2 Webshells

19.5.4.2.1 Manage webshells

This topic describes how to view and quarantine webshell files.

Context

Server Guard scans the Web directory on your server to check whether any webshell file exists. If a webshell file is detected, an alert is triggered.

Server Guard detects webshell files in PHP, JSP, or other common formats in real time or at scheduled time locally or in the cloud. Server Guard also allows you to quickly quarantine the detected webshell files.

Server Guard detects webshell files through the following methods:

- Dynamic detection: When any file in the Web directory is modified, Server Guard scans the modified content.
- Scheduled detection: Server Guard scans the entire Web directory every early morning.

> **Note:**
> By default, scheduled detection is enabled for all servers protected by Server Guard. To enable scheduled detection for a specific server, choose Settings > Security Settings. In the Trojan Scan area, click Manage on the right of Web

Directory Periodic Scan, and specify the server for which you want to enable scheduled detection.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Server Security > Intrusion Detection > Webshells.**
3. **Select a server to check the detected webshell files.**
4. **Handle the webshell files.**
   - **Click Quarantine to quarantine a file. You can select and quarantine multiple files at one time.**
   - **Click Restore to restore a file that has been quarantined by mistake.**
   - **Click Ignore to ignore a file. Server Guard no longer generates alerts for an ignored file.**

Note:
Server Guard does not delete webshell files on your server. Instead, it quarantines the files. You can restore a quarantined file if you determine that the file is trusted. After a webshell file is restored, Server Guard no longer generates alerts for this file.

19.5.4.3 Suspicious servers

19.5.4.3.1 Manage server exceptions

This topic describes how to view the alerts for server exceptions and handle the exceptions.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Server Security > Intrusion Detection > Suspicious Servers.**
3. **Select a server to view the detected exceptions.**
4. **Select an action to handle each exception based on its impact.**

Action	Description
Handle	Select this option to fix the exception.
Ignore Once	Select this option to ignore the alert if the exception does not have any impact on the server security.

Action	Description
Confirm	Select this option to confirm the exception.
Label as False Positive	Select this option if the alert is a false positive.
View	Select this option to view the alert details.

19.5.5 Server fingerprints

19.5.5.1 Manage listening ports

Security Center regularly collects information about listening ports on a server.

Context

This task is applicable to the following scenarios:

- **Check for servers that listen to the specified port.**
- **Check for the listening ports of a server.**

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Server Security > Server Security, and click the Listening Ports tab.**
3. **View the listening ports, network protocol, and the number of servers.**

   You can search for a port by the port number or process name.
4. **Click a port number to view the details, such as the corresponding asset and protocol.**

19.5.5.2 Manage processes

Security Center regularly collects the process information on a server.

Context

This task is applicable to the following scenarios:

- **Check for servers that run the specified process.**
- **Check for processes running on a server.**

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Server Security > Server Fingerprints, and click the Processes tab.**
3. **View all running processes and the number of servers that run these processes.**

   You can search for a process by process name or user.

4. Click a process name to view the details, such as the corresponding assets, path, and startup parameters.

19.5.5.3 Manage account information

Security Center regularly collects the account information on a server.

Context

This task is applicable to the following scenarios:

- Check for servers where the specified account is created.
- Check the accounts created on a server.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Server Security > Server Fingerprints, and click the Accounts tab.
3. View all accounts that have logged on and the number of servers that use these accounts.

   You can search for an account by account name.
4. Click an account name to view the details, such as the corresponding assets, root permissions, and user group.

19.5.5.4 Manage software versions

Security Center regularly collects software version information of a server.

Context

This task is applicable to the following scenarios:

- Check for software that has been installed without authorization.
- Check for software of outdated versions.
- Quickly locate the affected assets when vulnerabilities are detected.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Server Security > Server Fingerprints, and click the Software tab.
3. View all software in use and the number of servers that use such software.

   You can search by software name, version, or installation directory.
4. Click a software name to view the corresponding assets, software version, and other information.

19.5.5.5 Set the server fingerprint refresh frequency

You can set the frequency at which the data of running processes, system accounts, listening ports, and software versions is collected and refreshed.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Server Security > Server Fingerprints, and click Settings in the upper-right corner.**
3. **Select the refresh frequency from each drop-down list.**
4. **Click OK.**

19.5.6 Log retrieval

19.5.6.1 Log retrieval overview

The log retrieval function provided by Server Security allows you to manage logs scattered in various systems of Apsara Stack in a centralized manner, so that you can easily identify the causes of issues that occur on your servers.

The log retrieval function supports storage of logs for 180 days and query of logs generated within 30 days.

Benefits

The log retrieval function provides the following benefits:

- **End-to-end log retrieval platform: Allows you to retrieve logs of various Apsara Stack services in a centralized manner and trace issues easily.**
- **Cloud-based SaaS service: Allows you to query logs on all servers in Apsara Stack without additional installment and deployment.**
- **Supports TB-level data retrieval. It also allows you to add a maximum of 50 inference rules (Boolean expressions) in a search condition and obtain full-text search results within several seconds.**
- **Supports a wide range of log sources.**
- **Supports log shipping, which allows you to import security logs to Log Service for further analysis.**

Scenarios

You can use log retrieval to meet the following requirements:

- **Security event analysis: When a security event is detected on a server, you can retrieve the logs to identify the cause and assess the damage and affected assets.**
- **Operation audit: You can audit the operation logs on a server to identify high-risk operations and serious issues in a meticulous way.**

Supported log types

Table 19-17: Log types

Log type	Description
Logon history	Log entries about successful system logons
Brute-force attack	Log entries about system logon failures that are generated during brute-force attacks
Process snapshot	Log entries about processes on a server at a specific time
Listening port snapshot	Log entries about listening ports on a server at a specific time
Account snapshot	Log entries about account logon information on a server at a specific time
Process initiation	Log entries about process initiation on a server
Network connection	Log entries about active connections from a server to external networks

19.5.6.2 Log retrieval

This topic describes how to search for and view server logs.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Server Security > Log Retrieval.**
3. **Set search conditions.**

Table 19-18: Search condition parameters

Parameter	Description
Log source	Select a supported log source. For more information, see *Table 19-19: Log sources*.
Field	Select a field that is supported by the specified log source. For more information, see *Table 19-19: Log sources*.

Parameter	Description
Keyword	Enter the keyword of the field to be searched for.
Logical operator	Select a logical operator from value options: AND, OR, and NOT. For more information, see *Table 19-27: Logical operators*.
+	Add inference rules in a search condition for a log source.
Add conditions	Add search conditions for different log sources.

4. Click Search and view the search result.
   - Reset: Click Reset to clear the search condition configuration.
   - Save Search: Click Save Search to save the search condition configuration for future use.
   - Saved Searches: Click Saved Searches to select and apply a search condition configuration that has been saved.

19.5.6.3 Supported log sources and fields

This topic describes the types of logs and fields that are supported by the log retrieval function.

The log retrieval function allows you to query the following types of logs. You can click a log source to view the fields that can be retrieved.

Table 19-19: Log sources

Log source	Description
Logon log	Log entries about successful system logons.
Brute-force attack log	Log entries about system logon failures that are generated during brute-force attacks.
Process snapshot log	Log entries about processes on a server at a specific time.
Listening port snapshot log	Log entries about listening ports on a server at a specific time.
Account snapshot log	Log entries about account logon information on a server at a specific time.
Process initiation log	Log entries about process initiation on a server.

Log source	Description
Network connection log	**Log entries about active connections from a server to external networks.**

Logon log

The following fields are supported in logon log queries:

Table 19-20: Supported fields in logon logs

Field	Data type	Description
uuid	**string**	**The agent ID.**
IP	**string**	**The server IP address.**
warn_ip	**string**	**The source IP address for the logon.**
warn_port	**string**	**The logon port.**
warn_user	**string**	**The logon username.**
warn_type	**string**	**The logon type.**
warn_count	**string**	**The number of logon attempts.**
time	**datetime**	**The logon time.**

Brute-force attack log

The following fields are supported in brute-force attack log queries:

Table 19-21: Supported fields in brute-force attack logs

Field	Data type	Description
uuid	**string**	**The agent ID.**
IP	**string**	**The server IP address.**
warn_ip	**string**	**The attacker IP address.**
warn_port	**string**	**The target port number.**
warn_user	**string**	**The target username.**
warn_type	**string**	**The type.**
warn_count	**string**	**The number of brute-force attack attempts.**

Field	Data type	Description
time	datetime	The attack time.

Process initiation log

The following fields are supported in process initiation log queries.

Table 19-22: Supported fields in process initiation logs

Field	Data type	Description
uuid	string	The agent ID.
IP	string	The server IP address.
pid	string	The process ID.
groupname	string	The name of the user group.
ppid	string	The parent process ID.
uid	string	The user ID.
username	string	The username.
filename	string	The file name.
pfilename	string	The file name of the parent process.
cmdline	string	The command line.
filepath	string	The process path.
pfilepath	string	The parent process path.
time	datetime	The time when the process was started.

Listening port snapshot log

The following fields are supported in listening port snapshot log queries:

Table 19-23: Supported fields in listening port snapshot logs

Field	Data type	Description
uuid	string	The agent ID.
IP address	string	The server IP address.
src_port	string	The listening port.

Field	Data type	Description
src_ip	string	The listening IP address.
proc_path	string	The process path.
PID	string	The process ID.
proc_name	string	The process name.
proto	string	The protocol.
time	datetime	The time when data was collected.

Account snapshot log

The following fields are supported in account snapshot log queries:

Table 19-24: Supported fields in account snapshot logs

Field	Data type	Description
uuid	string	The agent ID.
IP address	string	The server IP address.
perm	string	Indicates whether the client has root permissions.
home_dir	string	The home directory.
warn_time	string	The password expiration notification time.
groups	string	The group to which the user belongs.
login_ip	string	The IP address of the last logon.
last_chg	string	The last time when the password was changed.
shell	string	The Linux shell command.
domain	string	The Windows domain.
tty	string	The logon terminal.
account_expire	string	The account expiration time.

Field	Data type	Description
passwd_expire	string	The password expiration time.
last_logon	string	The last logon time.
user	string	The user.
status	string	The user status. Value options include: · 0: disabled. · 1: normal.
time	datetime	The time when data was collected.

Process snapshot log

The following fields are supported in process snapshot log queries.

Table 19-25: Supported fields in process snapshot logs

Field	Data type	Description
uuid	string	The agent ID.
IP address	string	The server IP address.
path	string	The process path.
start_time	string	The time when the process was started.
uid	string	The user ID.
cmdline	string	The command line.
pname	string	The parent process name.
name	string	The process name.
pid	string	The process ID.
user	string	The username.
md5	string	The MD5 value of the process file. This value is not calculated if the file size exceeds 1 MB.
time	datetime	The time when data was collected.

Network connection log

The following fields are supported in network connection log queries.

Table 19-26: Supported fields in network connection logs

Field	Data type	Description
uuid	string	The agent ID.
IP	string	The server IP address.
src_ip	string	The source IP address.
src_port	string	The source port.
proc_path	string	The process path.
dst_port	string	The destination port.
proc_name	string	The process name.
dst_ip	string	The destination IP address .
status	string	The status.
proto	string	The protocol.
time	datetime	The connection time.

19.5.6.4 Inference rules and logical operators

The log retrieval function supports multiple search conditions. You can add multiple inference rules in one search condition for one log source, or combine multiple conditions for several log sources by using different logical operators. This topic describes the inference rules and logical operators that are supported in log queries. Some examples are provided to help you understand them.

The following table describes the logical operators that are supported in log queries.

Table 19-27: Logical operators

Logical operator	Description
AND	Binary operator. This operator is in the format of `query1 and query2`, indicating the intersection of the query results of `query1` and `query2`. **Note:** If no logical operator is specified for multiple keywords, the default operator is AND.
OR	Binary operator. This operator is in the format of `query1 or query2`, indicating the combination of the query results of `query1` and `query2`.
NOT	Binary operator. This operator is in the format of `query1 not query2`, indicating the results that match `query1` but not `query2`, which is equivalent to `query1 - query2`. **Note:** If only `not query1` is specified, the query returns all records that do not match `query1`.

19.5.7 Settings

19.5.7.1 Manage security settings

This topic describes how to manage the security settings of servers. You can enable or disable periodic trojan scan and set the resource usage mode of the Server Guard agent for servers.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Server Security > Settings.**

3. Enable periodic trojan scan for servers.
   a) Click Manage.
   b) Select the servers that require periodic trojan scan.
   c) Click OK.
4. Specify the resource usage mode of the Server Guard agent for servers.
   - Business First Mode: The peak CPU usage is less than 10% and the peak memory usage is less than 50 MB.
   - Protection First Mode: The peak CPU usage is less than 20% and the peak memory usage is less than 80 MB.

   a) Click Manage.
   b) Specify the work mode of the Server Guard agent for servers.
   c) Click OK.

19.5.7.2 Install the Server Guard agent

This topic describes how to manually install the Server Guard agent on a Windows or Linux server.

Prerequisites

If you have installed security software, such as Safedog and Yunsuo, on your server, the system may fail to install the Server Guard agent. We recommend that you disable or uninstall the security software before you install the agent.

Context

The Server Guard agent has been integrated in public images. If you select the public image when you create an ECS instance, the Server Guard agent is automatically integrated in the ECS instance.

For an external server that runs Windows, you must use the Server Guard agent installation package to install the agent. For an external server that runs Linux, you must run the relevant commands to install the agent.

To ensure that the agent can run correctly in the following situations, you must delete the Server Guard agent directory and use the preceding methods to manually reinstall the agent:

- An image that includes the Server Guard agent is used to install the agent on multiple external servers at one time.

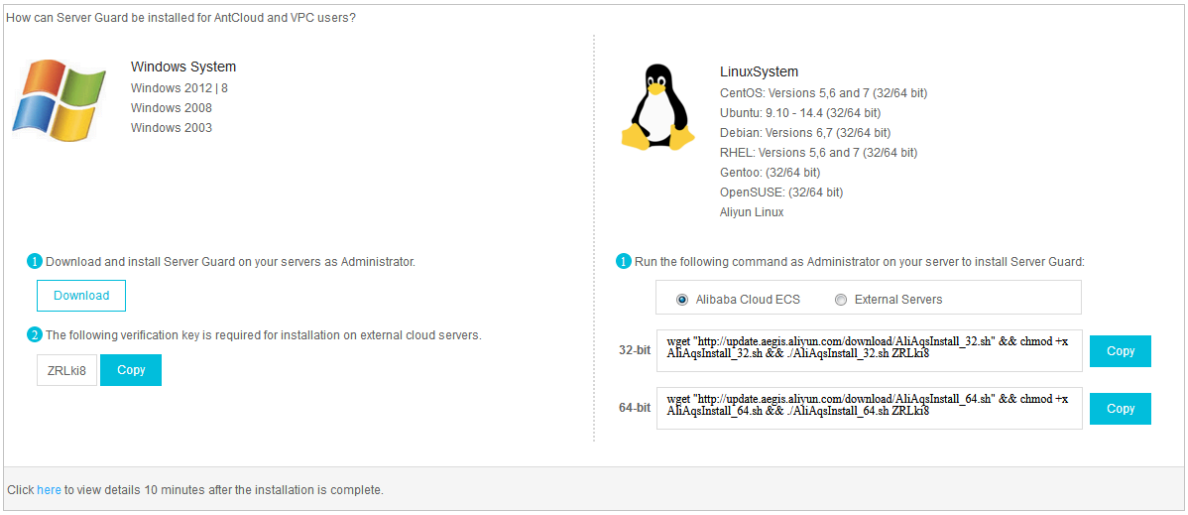
- You have copied the Server Guard agent files from a server that has been installed with the agent to your external servers.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Server Security > Settings > Install/Uninstall.

   The Server Guard agent installation page appears, as shown in *Figure 19-12: Install the agent*.

   Figure 19-12: Install the agent



3. Obtain and install the Server Guard agent based on the operating system type of your server.
   - Windows OS
     a. In the lower-left area of the page, click Download to download the installation package to your local PC.
     b. Upload the installation package to your server. For example, you can use an FTP client to upload the package to the server.
     c. Run the installation package on your server as an administrator.

        Note:

When installing the agent on an external server, you will be prompted to enter the installation verification key. You can find the installation verification key on the Server Guard agent installation page.

- Linux OS

  a. In the lower-right area of the page, select `Alibaba Cloud ECS` or `External Servers`.

  b. Select the installation command for your 32-bit or 64-bit operating system, and click Copy to copy the command.

  c. Log on to your Linux server as an administrator.

  d. Run the installation command on your Linux server to download and install the Server Guard agent.

4. View the agent status of your server.

   You can view the agent status of your server in the Server Guard console five minutes after you install the Server Guard agent.

   - If your server is an ECS instance, the status of the server changes from offline to online.
   - If your server is an external server, the server is added to the server list.

19.5.7.3 Uninstall the Server Guard agent from a server

If you decide not to use any of the Server Guard features on your server, you can use the following procedure to uninstall the Server Guard agent.

Context

Before you uninstall the Server Guard agent from a server in the console, make sure that the agent status of the server is online. If the status is offline, the server cannot receive the command for uninstalling the agent.

If you need to reinstall the Server Guard agent within 24 hours (the protection period) after the uninstallation, install it manually and ignore the error messages. You must run the install command at least three times before it can be successfully reinstalled.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Server Security > Settings > Install/Uninstall.
3. Click Uninstall in the upper-right corner.

4. In the Uninstall Server Guard dialog box, select the server from which you want to uninstall the Server Guard agent.
5. Click Uninstall. Then, the system automatically uninstalls the Server Guard agent.

19.6 Application security

19.6.1 Quick start

This topic helps you get started with Web Application Firewall (WAF).

WAF uses intelligent semantic analysis algorithms to identify Web attacks. WAF also integrates a learning model to enhance its analysis capabilities and meet your daily security protection requirements without relying on traditional rule libraries.

The procedure for using WAF is as follows:

1. Customize WAF protection rules.

   WAF provides a default protection policy. You can also customize policies that suit your business needs.

   - For more information about how to configure protection policies, see *Configure protection policies*.
   - For more information about how to configure custom rules, see *Create a custom rule*.
   - For more information about how to configure HTTP flood protection rules, see *Configure an HTTP flood protection rule*.

2. Add protected websites.

   WAF can protect Internet websites and Virtual Private Cloud (VPC) websites.

   - For more information about how to add an Internet website to WAF for protection, see *Add an Internet website for protection*.
   - For more information about how to add a VPC website to WAF for protection, see *Add a VPC website for protection*.

3. Configure Domain Name System (DNS) resolution.

   For more information about how to change the DNS-resolved source IP address of a website to a virtual IP address of WAF, see *Modify DNS resolution settings*.

4. **View WAF protection results.**
   - **For more information about how to view the overall protection information, see** *View protection overview*.
   - **For more information about how to view the access status, see** *View Web service access information*.
   - **For more information about how to view the detection logs for Web attacks, see** *View attack detection logs*.
   - **For more information about how to view the detection logs for HTTP flood attacks, see** *View HTTP flood protection logs*.

19.6.2 Protection configuration

19.6.2.1 Configure protection policies

This topic describes how to configure Web Application Firewall (WAF) protection policies.

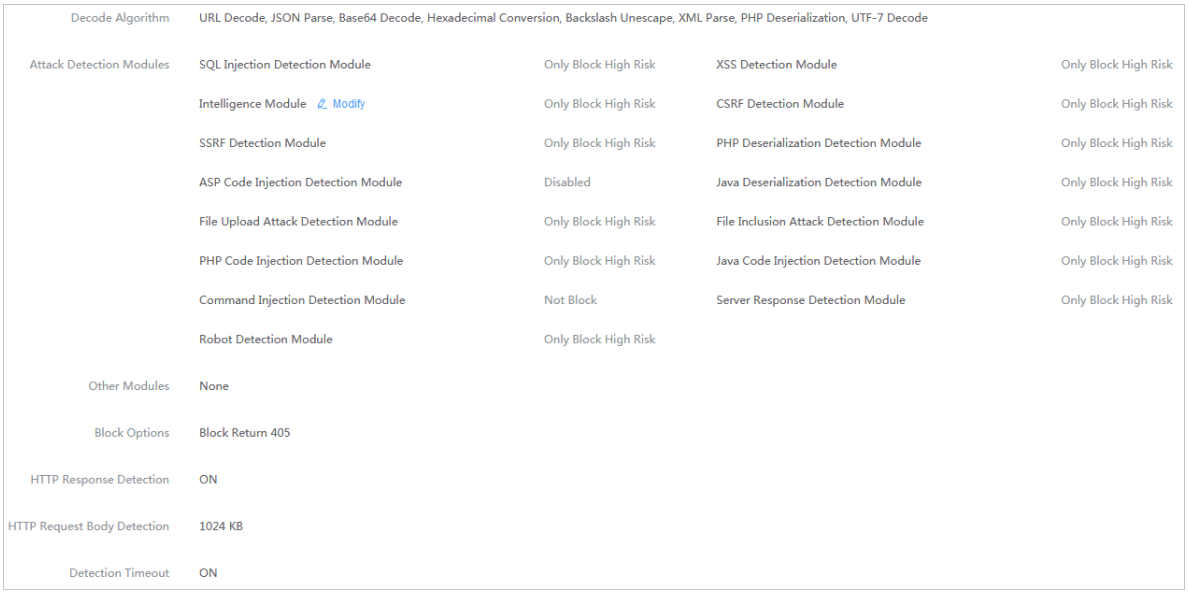
Context

WAF provides a default protection policy. You can also customize policies that suit your business needs.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Application Security > WAF. Choose Protection Configuration > Website Protection Policies.**
3. **Click Add Protection Policy. In the dialog box that appears, specify Policy Name, and click Confirm.**

4. Click the name of the new policy to modify the policy.



Parameter	Description
Decode Algorithm	Select algorithms for decoding the requests.
Attack Detection Modules	Specify the types of attacks to be detected and the risk levels of attacks to be blocked.
Block Options	Specify the status code and image to be returned when an attack is blocked.
HTTP Response Detection	Set the status of the Enable HTTP Response Detection toggle and specify Response Detection Max Body Size.
HTTP Request Body Detection	Specify Response Detection Max Body Size.

Parameter	Description
Detection Timeout	Set the status of the Enable Detection Timeout toggle, and specify Timeout Threshold.

For example, take the following steps to configure Attack Detection Modules:

a) Place the pointer over a specific module in the Attack Detection Modules area, for example, SQL Injection Detection Module, and click Modify.

b) In the SQL Injection Detection Module dialog box, set the detection parameters.

Parameter	Description
Enabled	Indicates whether to enable the detection module.
Blocking Threshold	You can select `Not Forbid`, `Only Forbid High Risk`, `Forbid Medium or High Risk`, or `Forbid All`.
Record Threshold	You can select `Not Record`, `Only Record High Risk`, `Record Medium or High Risk`, or `Record All`.
Detect Non-Injected SQL	Indicates whether to enable detection for non-injected SQL attacks.

c) Click Confirm.

5. **Manage the protection policies.**

   If you want to delete a protection policy, select this policy, and click Delete Selected Protection Policies.

   > **Note:**
   > **You cannot delete the default policy.**

19.6.2.2 Create a custom rule

This topic describes how to create a custom Web Application Firewall (WAF) rule.

Context

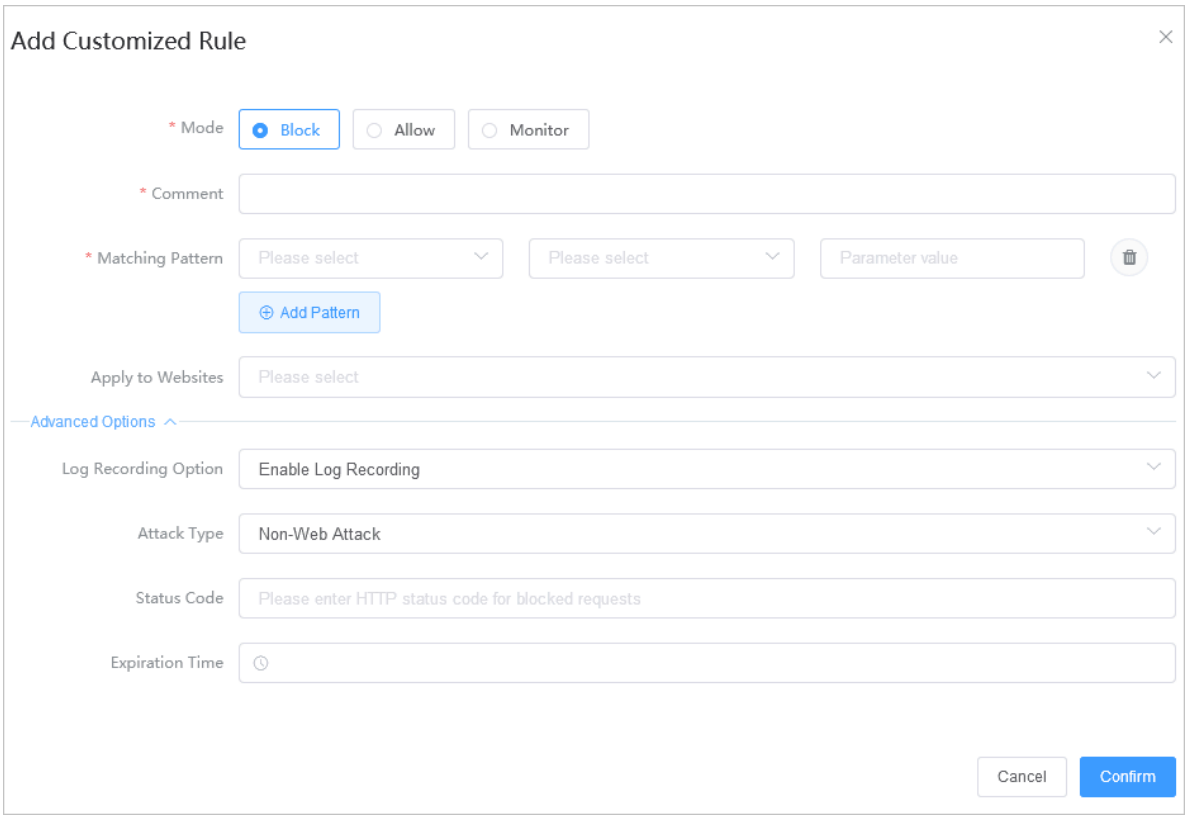
An enterprise administrator can create custom rules to meet various standards for attack detection. The administrator can add, edit, or delete the rules in Apsara Stack Console. A rule can be used to filter requests that meet certain conditions.

Multiple custom rules have logical OR relations. If two custom rules have the same conditions but differentiate in the mode such as blocking and allowing, the system runs the first rule.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. **Choose Application Security > WAF. Then, choose Protection Configuration > Customized Rules.**

3. Click Add New Customized Rule. In the Add Customized Rule dialog box, set the rule parameters.



Parameter	Description
Mode	The mode of the custom rule. Valid values: Block, Allow, and Monitor. · Block: Block an HTTP request if it meets the condition of this rule. · Allow: Allow an HTTP request if it meets the condition of this rule. · Monitor: Allow and monitor an HTTP request if it meets the condition of this rule.
Comment	The remarks for this rule, such as the purpose of this rule.

Parameter	Description
Matching Pattern	The conditions that trigger this rule. Some parameters in the conditions are described as follows: · URI: the requested URI before decoding. · Decoded path: Decode the part of the requested URL between the third slash (/) and the question mark (?) . For example, decode the part in bold in the following URL: `https://host:port/path?query`. · Query: Decode the part of the requested URL after the question mark (?) . For example, decode the part in bold in the following URL: `https://host:port/path?query`. · GET parameter: Convert the part of the requested URL after the question mark (?) to key-value pairs, and decode them to obtain the GET parameter. · Method: the method of the HTTP request. · Host: the host in the HTTP header. · Full cookie: the cookie in the HTTP header. Note: If you specify multiple conditions for one custom rule, these conditions have logical AND relations. The rule is triggered only when all conditions are met.
Apply to Websites	The websites to be protected by this rule.
Log Recording Option	Specifies whether to record a protection event in the attack detection logs if the rule is triggered. This option is enabled by default.
Attack Type	The type of attacks that are blocked by this rule.
Status Code	The status code to be returned after an attack is blocked by this rule.
Expiration Time	The time when this rule expires.

4. Click Confirm.

5. Manage custom rules.
   - Edit a rule.

     To edit a rule, click the Edit icon in the Actions column.
   - Enable a rule.

     To enable a rule, select this rule and click Enable Selected Rules.
   - Disable a rule.

     To disable a rule, select this rule and click Disable Selected Rules.
   - Delete a rule.

     To delete a rule, select this rule and click Delete Selected Rules.

19.6.2.3 Configure an HTTP flood protection rule

This topic describes how to configure an HTTP flood protection rule.

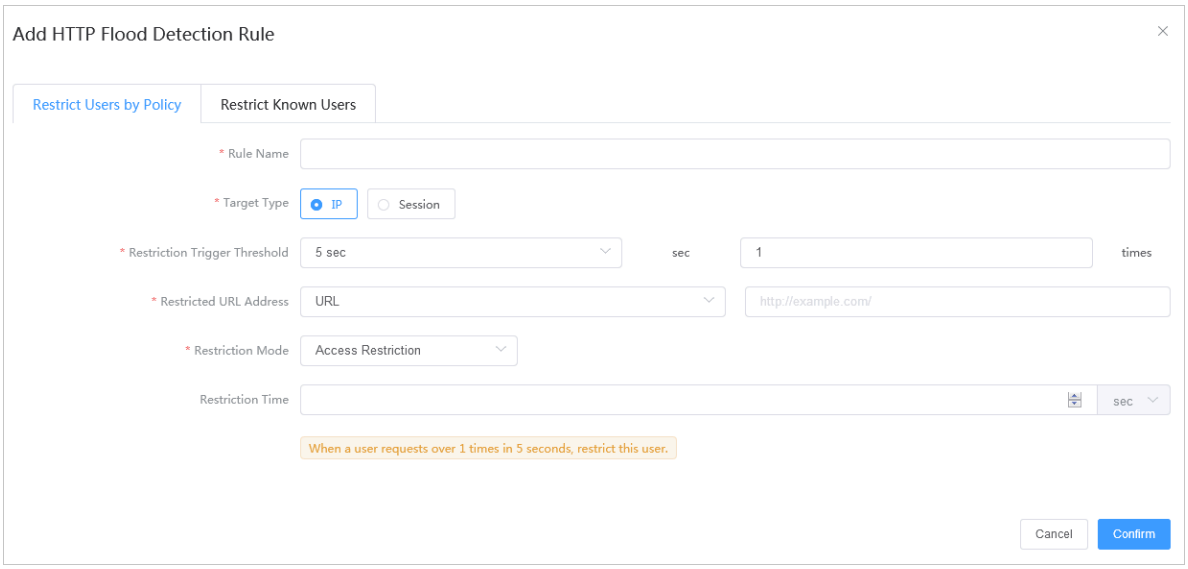
Context

An HTTP flood attack is a type of DDoS attack that is targeted at Web server applications. The attackers use proxy servers or bots to overwhelm a targeted Web server with HTTP requests.

Procedure

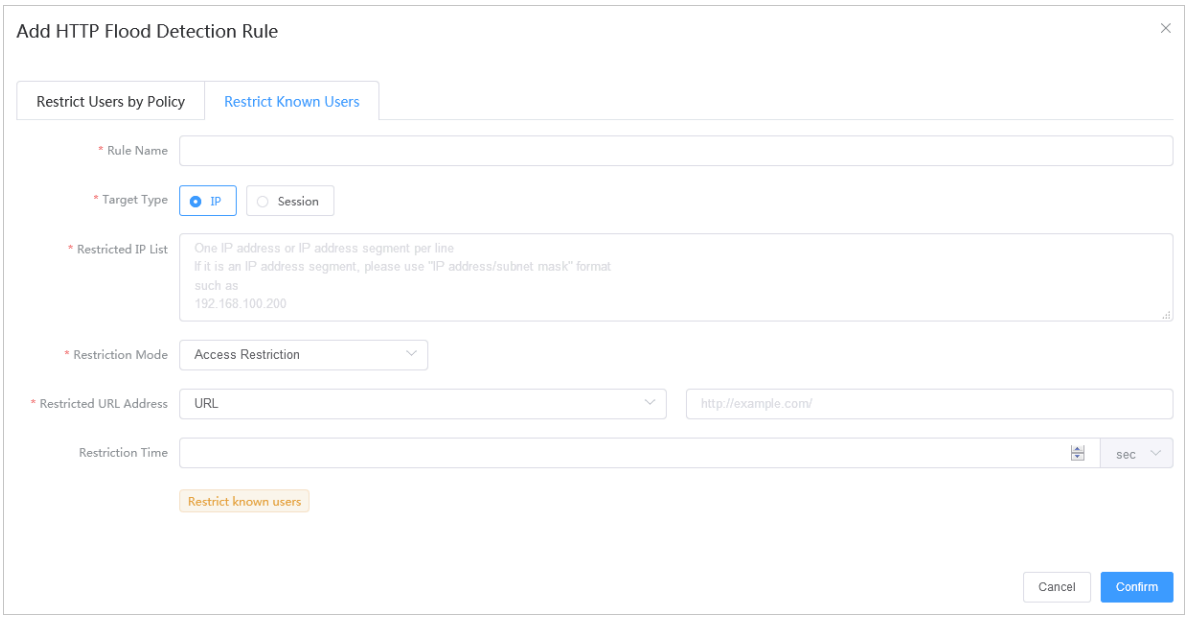
1. *Log on to Apsara Stack Security Center.*
2. Choose Application Security > WAF. Choose Protection Configuration > HTTP Flood Detection Rules.
3. Click Add Flood Detection Rule. The Add HTTP Flood Detection Rule page is displayed.
   - To restrict request sources that meet certain conditions, go to step *4*.
   - To restrict known request sources, go to *5*.

4. Click the Restrict Users by Policy tab, configure the rule parameters, and click Confirm.



Parameter	Description
Rule Name	Set the name of the HTTP flood protection rule.
Target Type	Set the type of the restricted target to `IP` or `Session`.
Restriction Trigger Threshold	Set the condition that triggers the restriction.
Restricted URL Address	Set the target URL address to be protected by this rule. • `URL` • `URL Prefix`
Restriction Mode	Set the mode in which the requests from a user is restricted. • `Access Restriction`: Forbid all requests from the restricted user to the specified URL. • `Access Frequency Restriction`: Limit the frequency of access from the restricted user to the specified URL.
Restriction Time	Set the time when the restriction takes effect.

5. Click the Restrict Known Users tab, configure the rule parameters, and click Confirm.



Parameter	Description
Rule Name	Set the name of the HTTP flood protection rule.
Target Type	Set the type of the restricted target to `IP` or `Session`.
Restricted IP List/Restricted Session List	Enter the IP addresses or sessions to be restricted based on the Target Type. Specify only one IP address or session in each line.
Restriction Mode	Set the mode in which the specified request source is restricted. • `Access Restriction`: Forbid all requests from the restricted user to the specified URL. • `Access Frequency Restriction`: Limit the frequency of access from the restricted user to the specified URL.
Restricted URL Address	Set the target URL address to be protected by this rule. • `URL` • `URL Prefix`
Restriction Time	Set the time when the restriction takes effect.

6. **Manage an HTTP flood protection rule.**
   - **Search for a rule.**

     Click Add Filter. Add filter conditions to quickly locate an HTTP flood protection rule.
   - **Enable a rule.**

     To enable a rule, select this rule, and click Enable Selected Rules.
   - **Disable a rule.**

     To disable a rule, select this rule, and click Disable Selected Rules.
   - **Delete a rule.**

     To delete a rule, select this rule, and click Delete Selected Rules.

19.6.2.4 Configure an HTTP flood protection whitelist

This topic describes how to configure an HTTP flood protection whitelist.

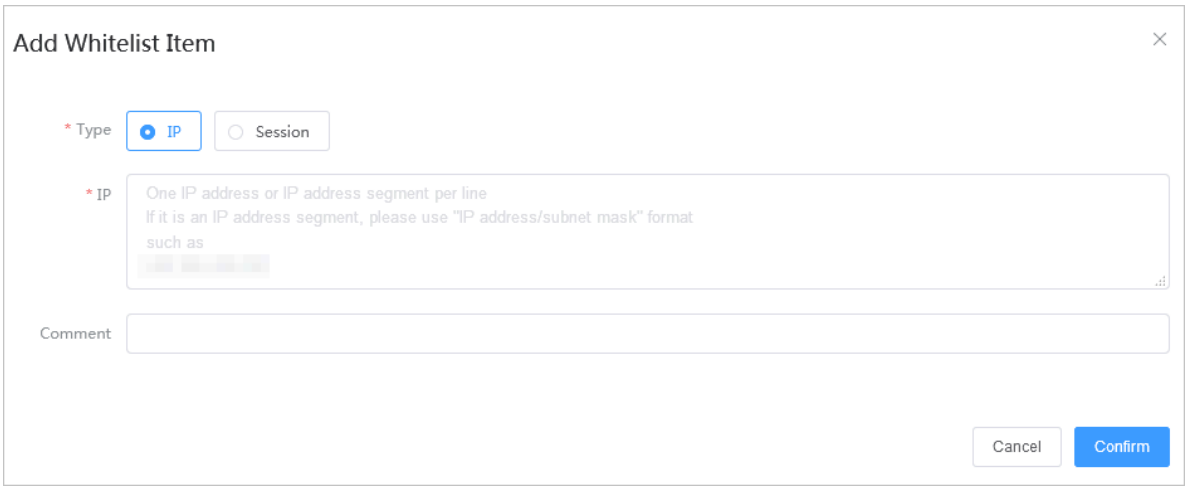
Context

If a request source is trusted, you can add this request source to an HTTP flood protection whitelist to allow the requests from this source.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Application Security > WAF. Choose Protection Configuration > HTTP Flood Detection Whitelist.**

3. Click Add Whitelist Item to add a request source to the whitelist, and click Confirm.



Parameter	Description
Type	Set the type of the whitelisted request source to `IP` or `Session`.
IP/Session	Specify the IP addresses or sessions based on the selected Type. Specify one IP address or session in each line.
Comment	Enter remarks for the whitelist.

4. Manage the whitelisted users.
   - Search for a whitelisted user.

     Click Add Filter. Add filter conditions to locate specific whitelisted users.
   - Remove a whitelisted user.

     To remove a request source from the whitelist, select the specific rule, and click Delete Selected Items.

19.6.2.5 Add an Internet website for protection

This topic describes how to add an Internet website to WAF for protection.

Context

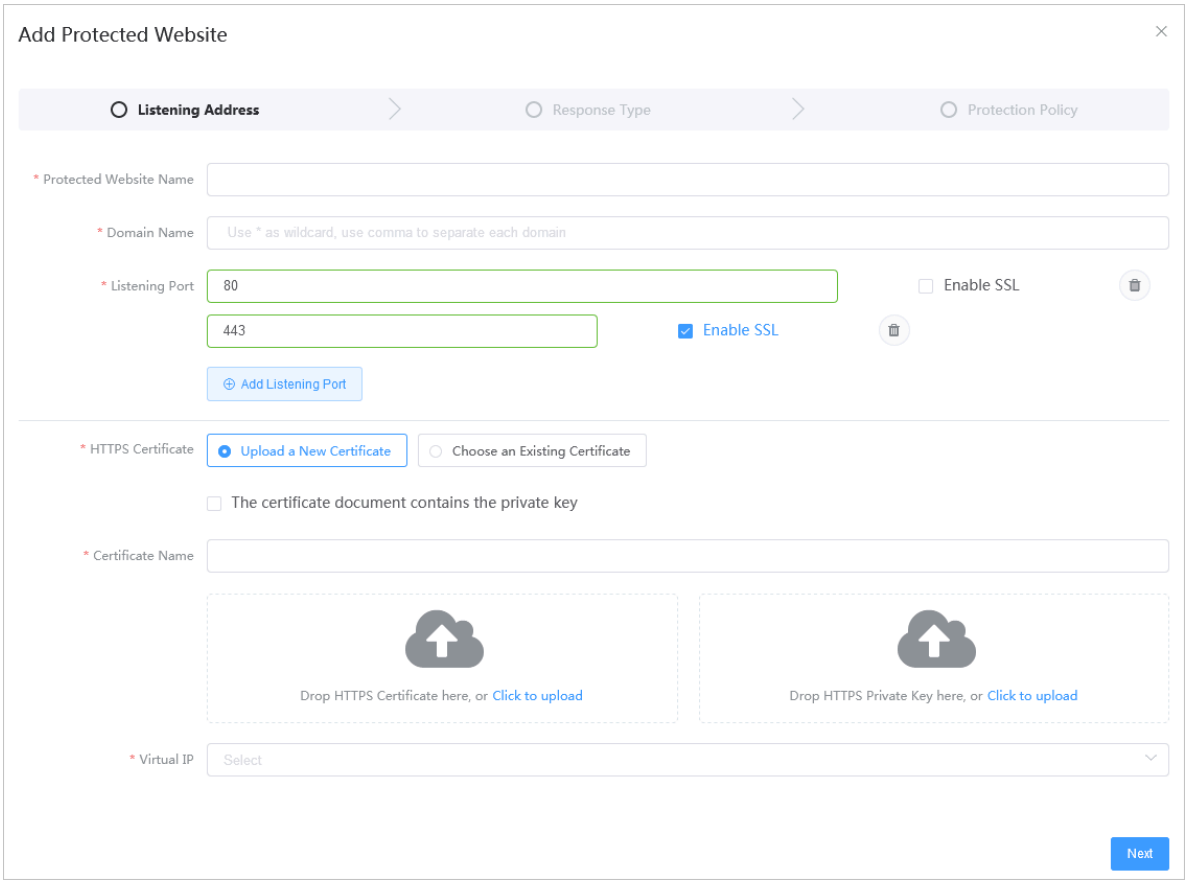
WAF can protect the following types of websites:

- Internet websites.
- Virtual Private Cloud (VPC) websites: For more information, see *Add a VPC website for protection*.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose Application Security > WAF > Protection Configuration > Protected Websites. On the page that appears, click the Internet Websites tab.
3. Click Add Protected Website.
4. In the Listening Address step, set the parameters and click Next.

   Set the Internet website to be protected. WAF can protect HTTP and HTTPS websites.

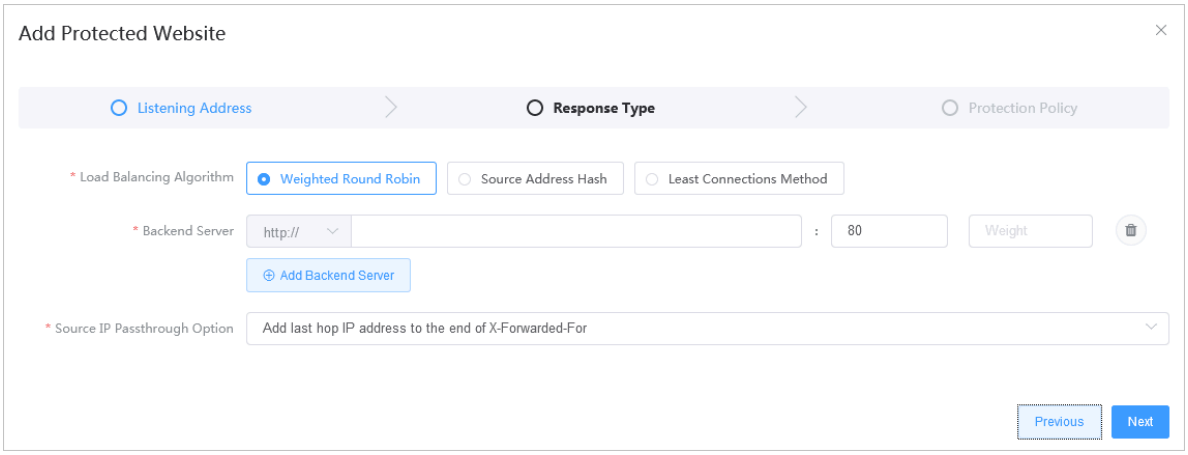


Parameter	Description
Protected Website Name	The name of the website to be protected.
Domain Name	The domain name of the website. · You can use an asterisk (*) as a wildcard. · Separate multiple domain names with commas (,).

Parameter	Description
Listening Port	The port listened by WAF. · If the website can be accessed through HTTPS, select the Enable SSL check box and upload the HTTPS certificate. · If the website can be accessed through multiple ports, click Add Listening Port to add a port.
HTTPS Certificate	The HTTPS certificate of the website. Valid values: `Upload a New Certificate` and `Choose an Existing Certificate`. · `Upload a New Certificate`: Select this option if the HTTPS certificate used by the website has not been uploaded to WAF before. By default, the HTTPS certificate and private key are uploaded separately. If you select The certificate document contains the private key, you only need to upload a file that contains the HTTPS certificate and private key. · `Choose an Existing Certificate`: If the HTTPS certificate used by the website has been uploaded to WAF before, select this option, and then select the HTTPS certificate from the drop-down list. Note: Set this parameter only when you select Enable SSL in the Listening Port step.
Certificate Name	The name of the HTTPS certificate. Note: Set this parameter only when you select Enable SSL in the Listening Port step.

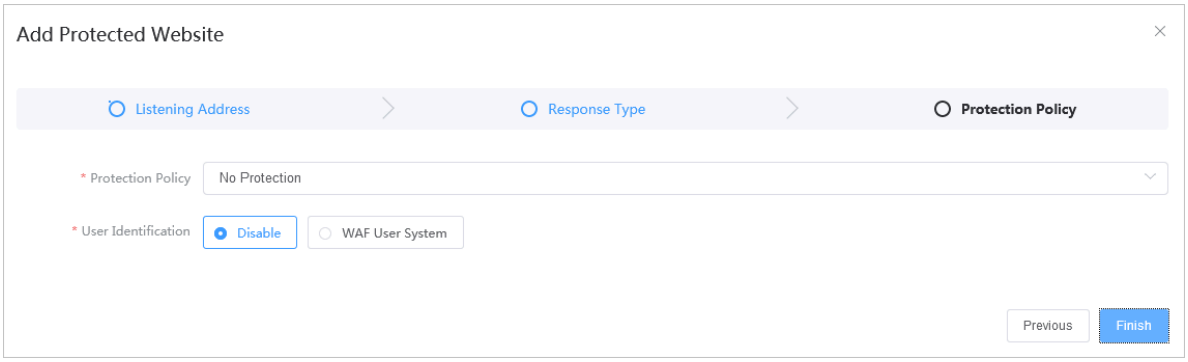
Parameter	Description
Areas for uploading the HTTPS certificate and private key	Upload the HTTPS certificate and private key. By default, the HTTPS certificate and private key are uploaded separately. If you select The certificate document contains the private key, you only need to upload a file that contains the HTTPS certificate and private key. **Note:** Upload the HTTPS certificate and private key only when you select Enable SSL in the Listening Port step.
Virtual IP	The virtual IP address of the website.

5. In the Response Type step, set the parameters and click Next.



Parameter	Description
Load Balancing Algorithm	The algorithm for balancing load. Valid values: `Weighted Round Robin`, `Source Address Hash`, and `Least Connections Method`.
Backend Server	The address of the back-end server.
Source IP Passthrough Option	The transparent transmission mode of the source IP address. The X-Forwarded-For (XFF) HTTP header field is a common method for identifying the original IP address of an HTTP client. It is used in request forwarding services such as HTTP proxy and load balancing.

6. In the Protection Policy step, set the parameters and click Finish.



Parameter	Description
Protection Policy	The WAF protection policy. For more information, see *Configure protection policies*.
User Identification	Specifies whether to enable the user identification feature.

19.6.2.6 Add a VPC website for protection

This topic describes how to add a Virtual Private Cloud (VPC) website to Web Application Firewall (WAF) for protection.

Context

WAF can protect the following types of websites:

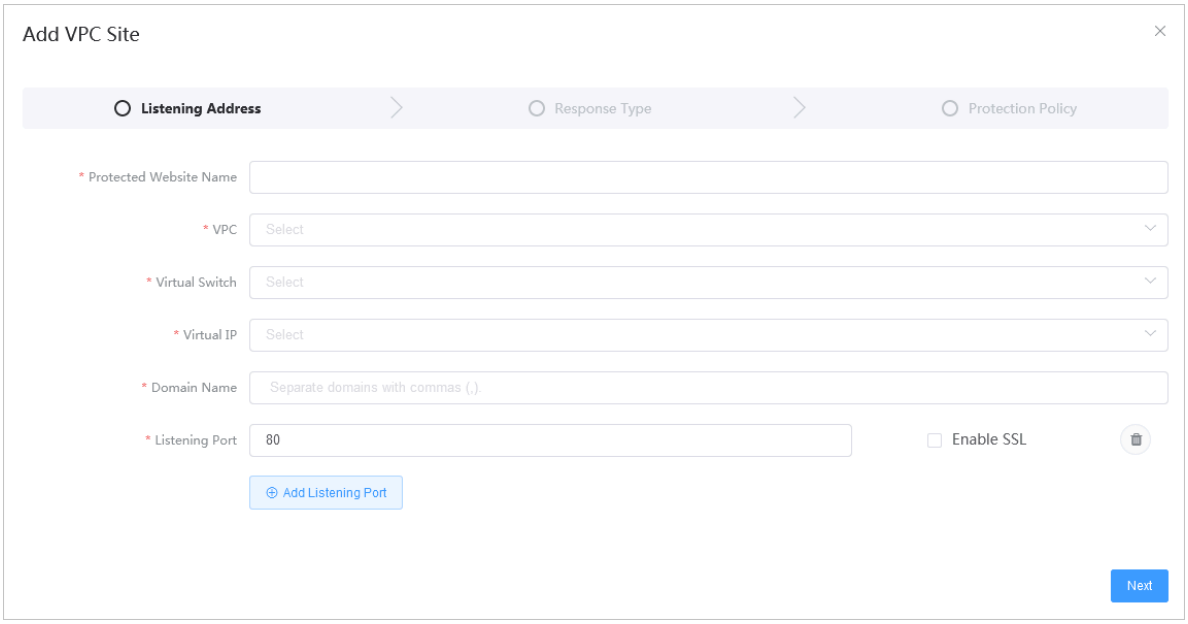
- Internet websites. For more information, see *Add an Internet website for protection*.
- VPC websites.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose Application Security > WAF > Protection Configuration > Protected Websites. On the page that appears, click the VPC Websites tab.
3. Click Add VPC Website. The Add VPC Site wizard appears.

4. In the Listening Address step, set the parameters and click Next.

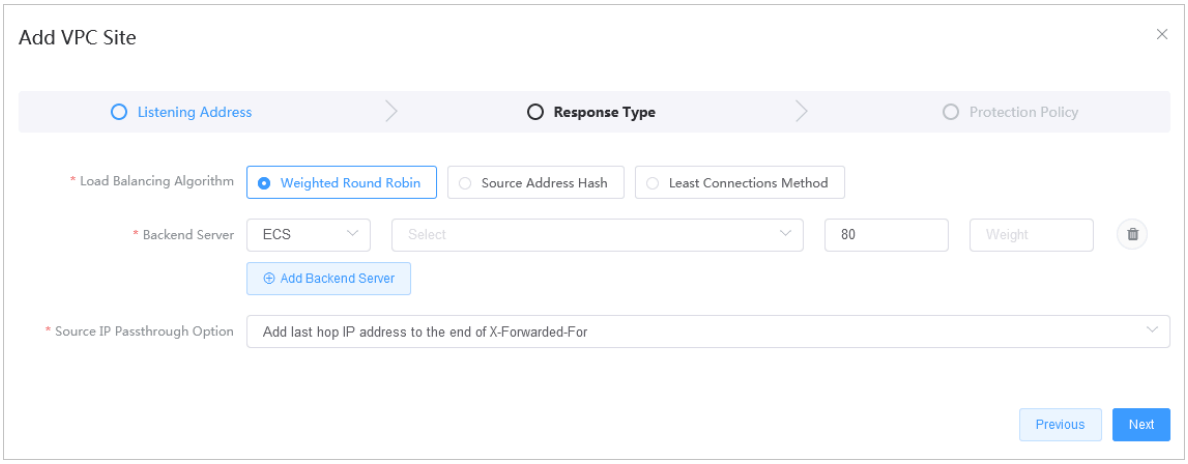
   Set the VPC website to be protected. WAF can protect HTTP and HTTPS websites.



Parameter	Description
Protected Website Name	The name of the website to be protected.
VPC	The VPC to which the website belongs.
Virtual Switch	The Vswitch to which the website belongs.
Virtual IP	The virtual IP address of the website.
Domain Name	The domain name of the website. · You can use an asterisk (*) as a wildcard. · Separate multiple domain names with commas (,).
Listening Port	The port listened by WAF. · If the website can be accessed through HTTPS, select the Enable SSL check box and upload the HTTPS certificate. · If the website can be accessed through multiple ports, click Add Listening Port to add a port.

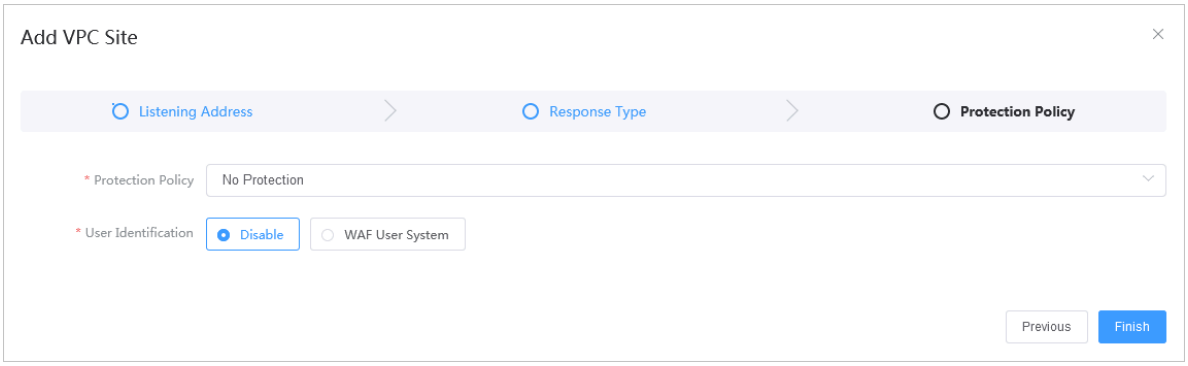
Parameter	Description
HTTPS Certificate	The HTTPS certificate of the website. Valid values: `Upload a New Certificate` and `Choose an Existing Certificate`. • `Upload a New Certificate`: Select this option if the HTTPS certificate used by the website has not been uploaded to WAF before. By default, the HTTPS certificate and private key are uploaded separately. If you select The certificate document contains the private key, you only need to upload a file that contains the HTTPS certificate and private key. • `Choose an Existing Certificate`: If the HTTPS certificate used by the website has been uploaded to WAF before, select this option, and then select the HTTPS certificate from the drop-down list. Note: Set this parameter only when you select Enable SSL in the Listening Port step.
Certificate Name	The name of the HTTPS certificate. Note: Set this parameter only when you select Enable SSL in the Listening Port step.
Areas for uploading the HTTPS certificate and private key	Upload the HTTPS certificate and private key. By default, the HTTPS certificate and private key are uploaded separately. If you select The certificate document contains the private key, you only need to upload a file that contains the HTTPS certificate and private key. Note: Upload the HTTPS certificate and private key only when you select Enable SSL in the Listening Port step.

5. In the Response Type step, set the parameters and click Next.



Parameter	Description
Load Balancing Algorithm	The algorithm for balancing load. Valid values: `Weighted Round Robin`, `Source Address Hash`, and `Least Connections Method`.
Backend Server	The address of the back-end server.
Source IP Passthrough Option	The transparent transmission mode of the source IP address. The X-Forwarded-For (XFF) HTTP header field is a common method for identifying the original IP address of an HTTP client. It is used in request forwarding services such as HTTP proxy and load balancing.

6. In the Protection Policy step, set the parameters and click Finish.



Parameter	Description
Protection Policy	The WAF protection policy. For more information, see *Configure protection policies*.
User Identification	Specifies whether to enable the user identification feature.

19.6.2.7 Verify the WAF connection configuration for a domain name locally

This topic describes how to verify the WAF connection configuration for a domain name by accessing the domain name from a local PC.

Context

Before you redirect business traffic to WAF, we recommend that you perform a local verification to ensure that the domain name has been connected to WAF and that WAF can forward traffic correctly. After you have added the virtual IP of WAF and the domain name of a website to the local hosts file, the request to access the domain name from a local browser passes through WAF first.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Add the virtual IP and domain name to the** `hosts` **file on your local PC.**

   For example, in Windows 7, the hosts file path is `C:\Windows\System32\drivers\etc\hosts`**.**

   a) **Open the hosts file by using a text editor such as Notepad.**

   b) **Add the following line at the end of the file:** `<Protected website virtual IP address><Protected website domain>`**.**

   ```
   # localhost name resolution is handled within DNS itself.
   #	127.0.0.1       localhost
   #	::1             localhost

   [redacted]4.115      example.com
   ```

   Note:
   The IP address in front of the domain name is the virtual IP address assigned by WAF.
3. **Ping the protected domain name from the local PC.**

   The resolved IP address must be the WAF virtual IP in the hosts file. If the resolved IP address is still the IP address of the origin website, refresh the local DNS cache.

4. Enter the domain name in the address bar of a browser and press Enter.

   If the domain name has been connected to WAF, you can visit the website.

5. Verify the WAF protection feature.

   Simulate a Web attack request to check whether WAF blocks the request.

   For example, add `/? alert(xss)` after the URL. If you try to visit `www.example.com /? alert(xss)`, WAF must block the request.

19.6.2.8 Modify DNS resolution settings

This topic describes how to connect your businesses to WAF by modifying the DNS resolution settings.

Context

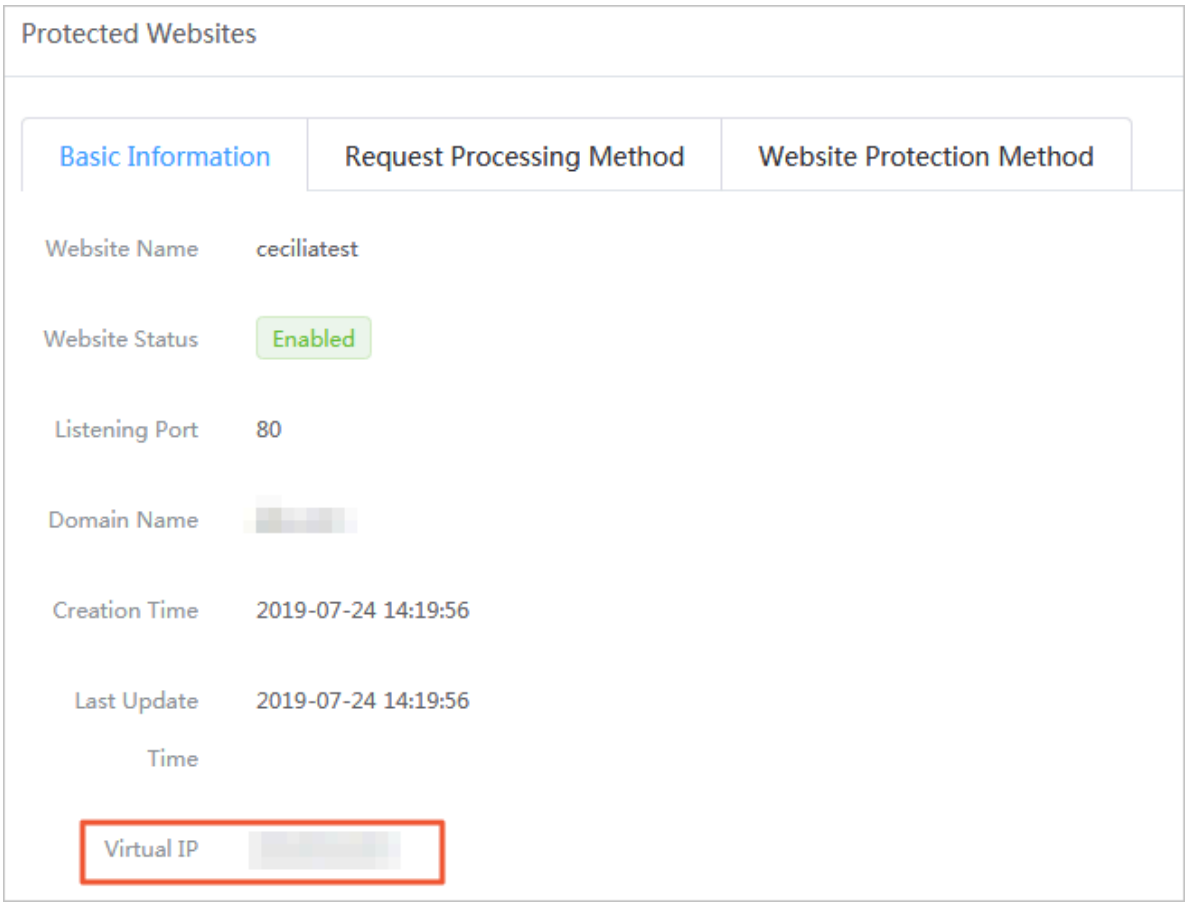
Before you modify the DNS resolution settings and redirect business traffic to WAF, make sure that you have passed local verification.

The domain name of a protected website may not be resolved by a DNS provider, for example, a website may use a Server Load Balancer (SLB) instance to connect to the Internet. To connect such a domain name to WAF, use the following procedure to specify the virtual IP address of the protected website as the origin IP address of the SLB instance:

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Application Security > WAF. Choose Protection Configuration > Protected Website List.
3. Click the name of a website.

4. On the Basic Information tab page, obtain the virtual IP address of the protected website.



5. Log on to the console provided by the DNS provider and find the domain name resolution settings for the relevant domain name. Then, change the A record value to the virtual IP address of the protected site.

> **Note:**
> We recommend that you set the TTL to 600 seconds in DNS resolution settings. The greater the TTL is, the longer it takes to synchronize and update DNS records.

19.6.3 Detection overview

19.6.3.1 View protection overview

This topic describes how to view the WAF protection overview.

Context

The Protection Overview page displays the statistics of previous attacks, the geographical distribution of attackers, the numbers of total requests and blocked

requests, and other information. You can quickly learn the Web attack protection information and custom protection rules.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Application Security > WAF. Choose Detection Overview > Detection Overview.**

3. On the Detection Overview page, you can view Statistics within Last 24 Hours and Statistics within Last 30 Days.

- Attacker geographical distribution on a map

  The distribution of attackers is displayed on a map. You can select a map of China or a map of the world.

  The numbers of total requests and blocked requests are displayed.
- Distribution of top five attack types

  A pie chart is provided to display the distribution of the top five attack types and the number of attacks of each type.
- Top five attacked websites

  A bar chart is provided to display the top five attacked websites and the number of attacks on each website.

19.6.3.2 View Web service access information

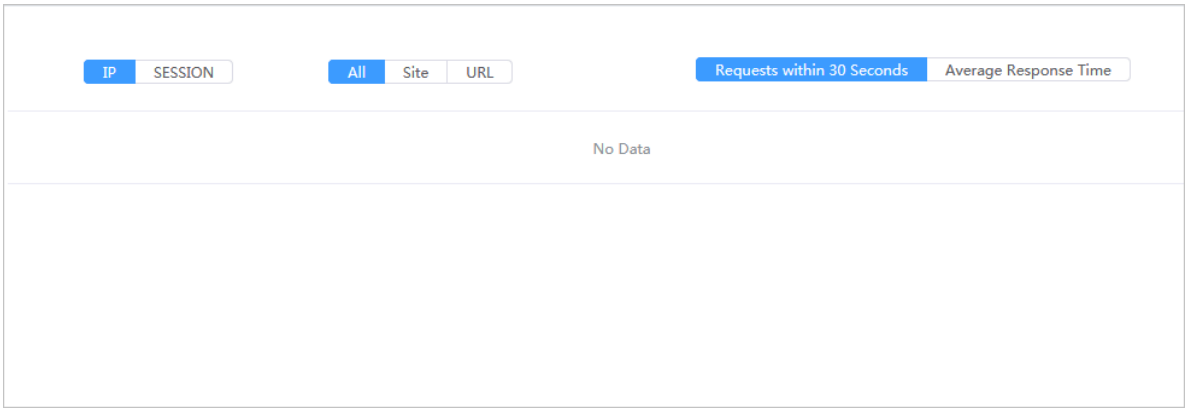
This topic describes how to view the service access information.

Context

WAF monitors the Web service access information. This allows security administrators to analyze the business access information and detect vulnerabilities.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Application Security > WAF. Choose Detection Overview > Access Status Monitor.**
3. **Filter the access records to view the details.**



19.6.4 Protection logs

19.6.4.1 View attack detection logs

This topic describes how to view attack detection logs.

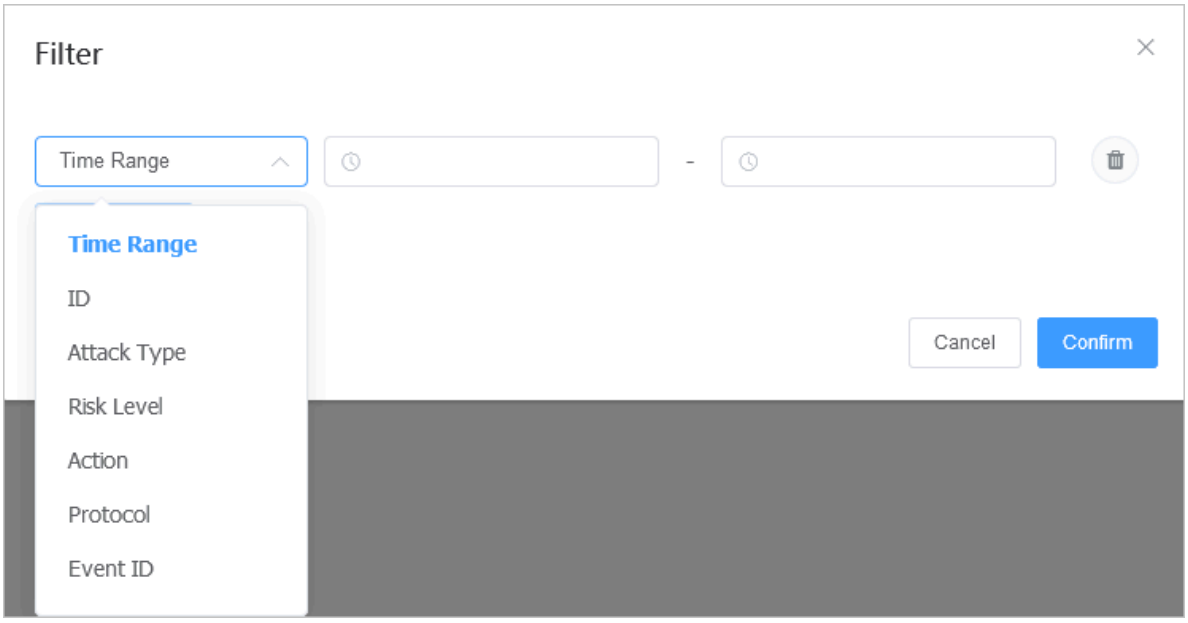
Context

These logs allow you to analyze the attacks on your Web services. Based on the analysis, you can update the attack protection policies and custom rules and fix the Web service vulnerabilities.

Procedure

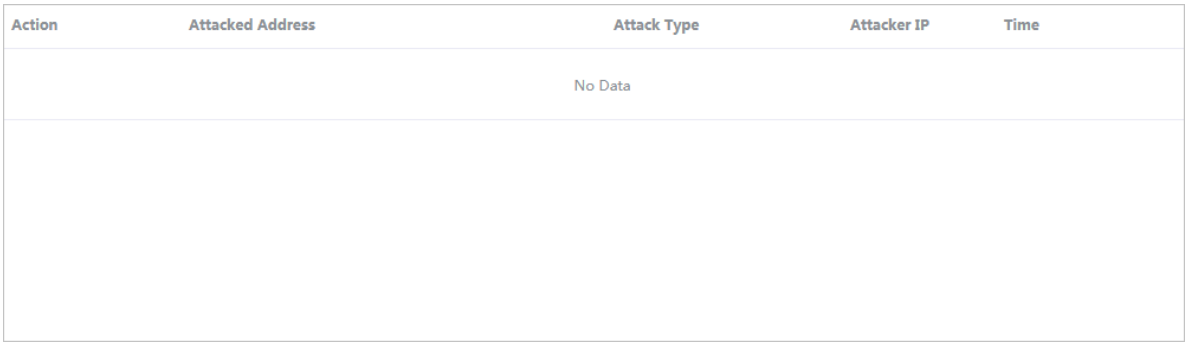
1. *Log on to Apsara Stack Security Center.*
2. **Choose Application Security > WAF. Choose Detection Logs > Attack Detection Logs.**

3. Click Add Filter, specify filter conditions, and click Confirm.



Note:
If you specify multiple conditions, all of the conditions must be met.

4. View the detected attacks.



19.6.4.2 View HTTP flood protection logs

This topic describes how to view HTTP flood protection logs.

Context

These logs allow you to analyze HTTP flood attacks on your Web services. Based on the analysis, you can update the HTTP flood protection rules and HTTP flood whitelist and fix the Web service vulnerabilities.

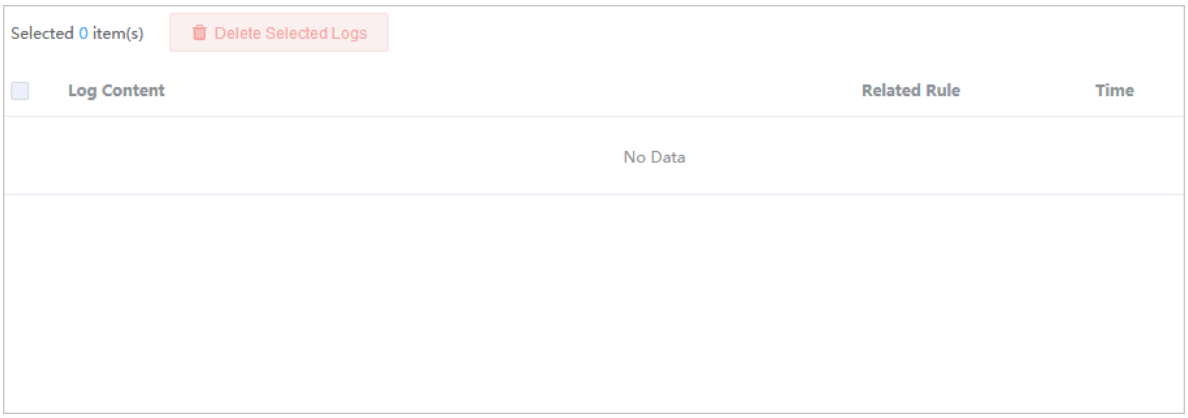
Procedure

1. *Log on to Apsara Stack Security Center.*

2. Choose Application Security > WAF. Choose Detection Logs > HTTP Flood Detection Logs.
3. Click Add Filter, specify filter conditions, and click Confirm.

Note:
If you specify multiple conditions, all of the conditions must be met.

4. View the HTTP flood detection result.



The blocked HTTP flood attacks, related rules, and attack time are displayed.

You can select a log and click Delete Selected Logs to delete a log.

19.7 System management

19.7.1 Manage accounts

This topic describes how to manage your Apsara Stack account bound to Apsara Stack Security.

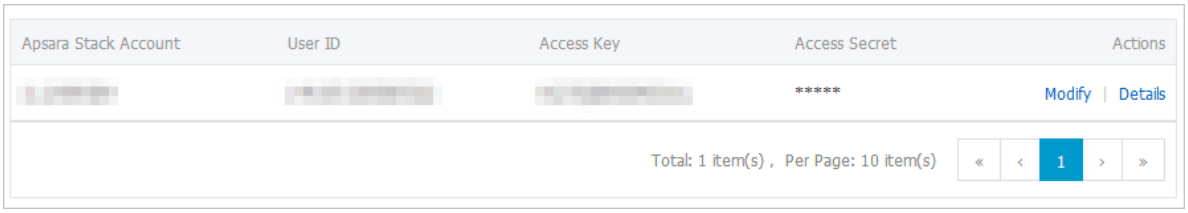
Context

Note:

All assets in Apsara Stack Security are bound to your Apsara Stack account. Exercise caution when you modify the account information.

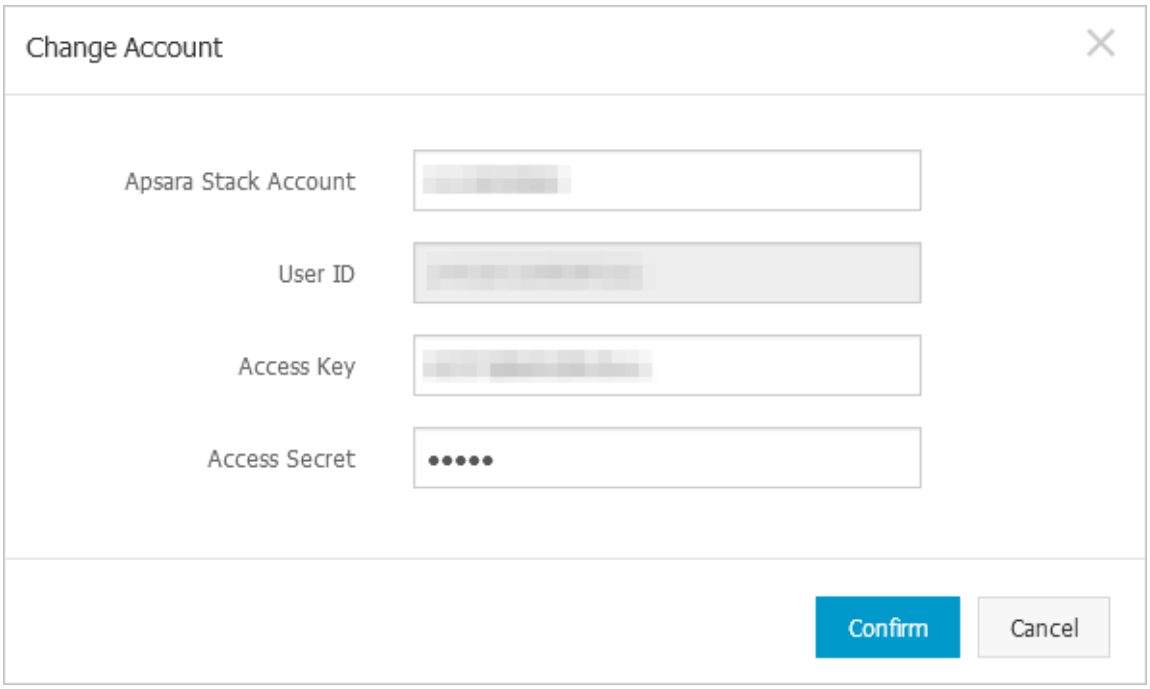
Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose System Management > Account Management.



You can view and modify your Apsara Stack account bound to Apsara Stack Security.

3. Modify the information about your Apsara Stack account.
   a) Click Change.
   b) In the Change Account dialog box that appears, modify the account information.



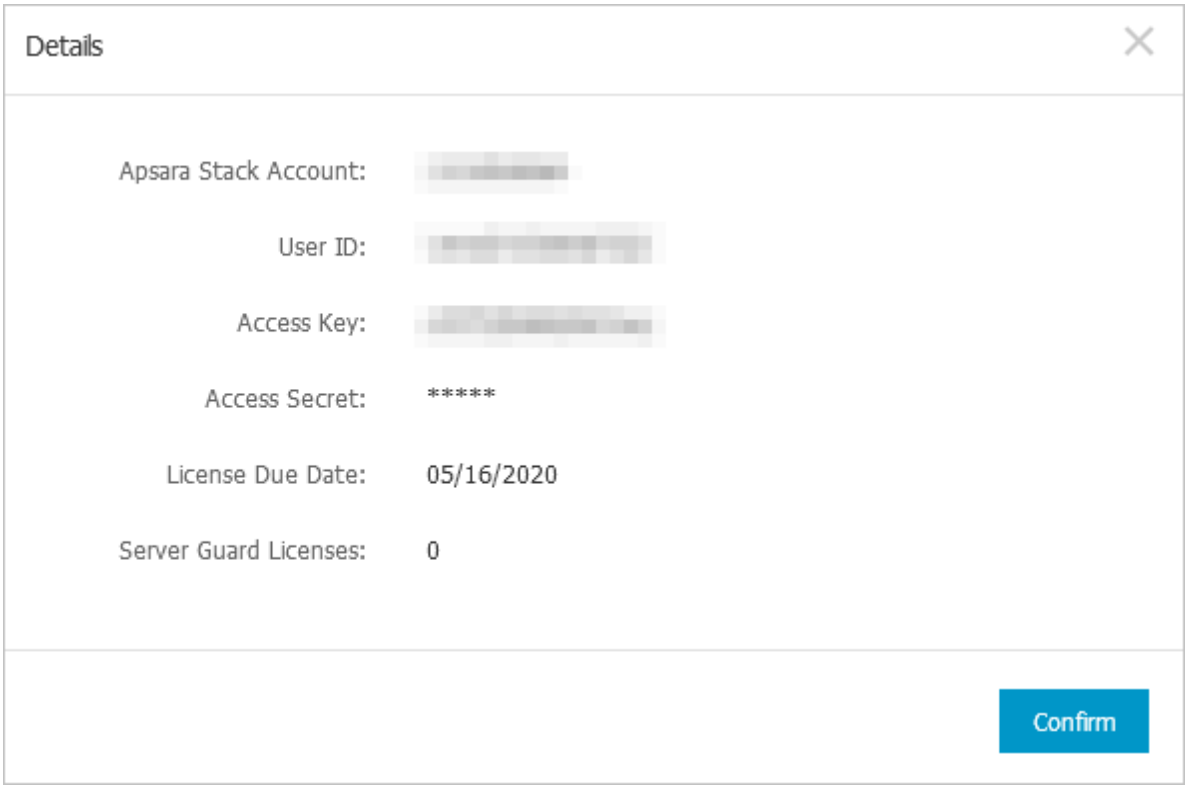
   c) Click OK.

4. View the details of your Apsara Stack account.

   a) Click Details.

   b) View the details of your Apsara Stack account.

   The details include the license expiration time and the number of licenses for Server Guard. The information is obtained based on your AccessKey.



19.7.2 Alert settings

19.7.2.1 Set alert recipients

This topic describes how to add and manage alert recipients.

Context

Alert recipients are those who receive alert notifications. Alert notifications can be sent by SMS or email. When the monitored data meets an alert rule, an alert notification is sent to the alert recipient.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose System Management > Alert Settings and then click the Alert Recipient tab.
3. Click Add Recipient.

4. Enter the recipient information and click OK.
5. Manage alert recipients.

   In the recipient list, click Edit or Delete in the Actions column of a recipient to edit or delete the recipient.

19.7.2.2 Set alert notifications

This topic describes how to set the method for sending alert notifications for various security events.

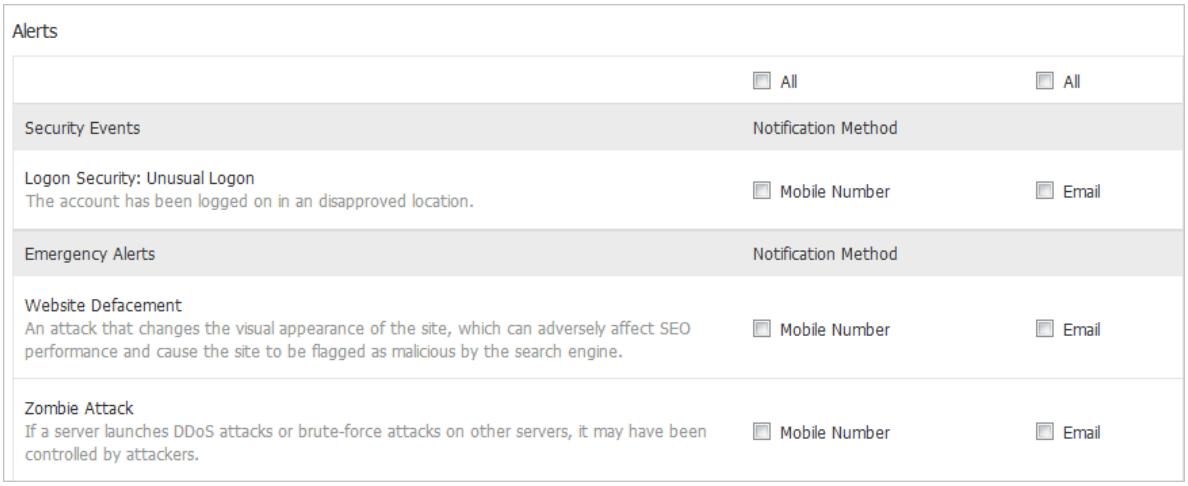
Context

In the Alerts section, the security administrator can set the method for sending alert notifications for various security events. When a security event occurs, the system notifies the alert recipients by email or SMS. For more information about how to set alert recipients, see *Set alert recipients*.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose System Management > Alert Settings and then click the Alert Settings tab.
3. In the Alerts section, select a notification method for each security event.

   Figure 19-13: Alert Settings page



4. Click Confirm.

19.7.3 Global settings

19.7.3.1 Set CIDR blocks for traffic monitoring

19.7.3.1.1 Add a CIDR block for traffic monitoring

This topic describes how to add a CIDR block for traffic monitoring. The traffic security monitoring module of Apsara Stack Security monitors the traffic of the specified CIDR block.

Context

CIDR blocks are configured for the traffic security monitoring module. The security administrator can change the CIDR blocks for monitoring as needed. The settings of CIDR blocks for traffic monitoring only apply to data centers in the correspond ing regions.

> **Note:**
>
> **Changes to CIDR block settings take effect immediately without the intervention of security administrators.**
>
> **If the same CIDR block is configured for the traffic security monitoring module and region detection, make sure that the same region is specified for the CIDR block.**

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose System Management > Global Settings > Traffic Collection IP Range.**
3. **Click Add.**

4. In the Add CIDR Block for Monitoring dialog box, specify a CIDR block.

   - CIDR Block: Enter a CIDR block for traffic monitoring.

     Note:
     Enter a valid CIDR block. You cannot enter a CIDR block that already exists in the system.

   - Region: Specify the region of the data center.

5. Click Confirm.

19.7.3.1.2 Manage CIDR blocks for traffic collection

This topic describes how to modify or delete Classless Inter-Domain Routing (CIDR) blocks for traffic collection.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose System Management > Global Settings > Traffic Collection IP Range.
3. Select a region, enter the target CIDR block, and click Search.

   View the information about the CIDR block for traffic collection and region in the search result.

4. Manage a CIDR block for traffic collection by clicking a button in the Actions column.
   - Modify the CIDR block for traffic collection.

     Click Modify to modify the region of the CIDR block for traffic collection.
   - Delete the CIDR block for traffic collection.

     Click Delete to delete the CIDR block for traffic collection.

19.7.3.2 Region settings

19.7.3.2.1 Add a CIDR block for a region

This topic describes how to add CIDR blocks for regions that are detected and reported by Server Guard.

Context

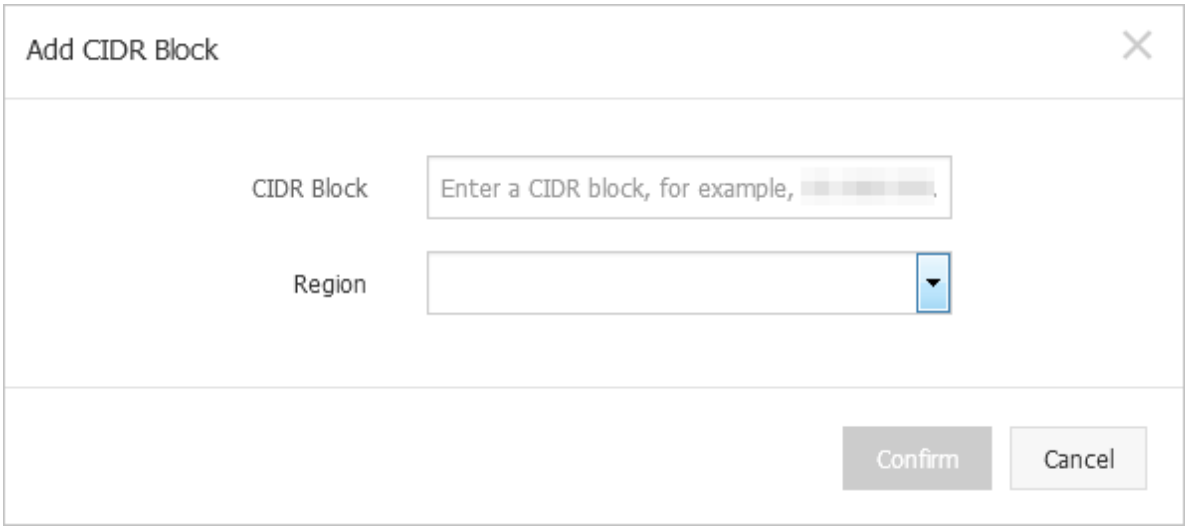
Region settings are used for region detection of Server Guard agents. Server Guard servers automatically detects and matches the regions of servers based on the IP address information reported by Server Guard agents.

> **Note:**
> **You can change the region of a CIDR block. After modification, you must modify the region for all assets in the CIDR block on the Asset Overview page.**

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose System Management > Global Settings > Region.
3. Click Add.

4. In the Add CIDR Block dialog box, set the CIDR block.



   - CIDR Block: Enter a CIDR block for the region.

     Note:
     Enter a valid CIDR block. You cannot enter a CIDR block that already exists in the system.

   - Region: Specify the region.

5. Click Confirm.

19.7.3.2.2 Manage CIDR blocks for a region

This topic describes how to modify or delete Classless Inter-Domain Routing (CIDR) blocks for a region.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose System Management > Global Settings > Region.
3. Select a region, enter the target CIDR block, and click Search.

   View the information about the CIDR block for the region in the search result.
4. Manage a CIDR block for the region by clicking a button in the Actions column.
   - Modify the CIDR block for the region.

     Click Modify to modify the CIDR block for the region.
   - Delete the CIDR block for the region.

     Click Delete to delete the CIDR block for the region.

19.7.3.3 Configure whitelists

This topic describes how to configure the brute-force attack blocking whitelist in Server Guard and the following whitelists in Threat Detection Service: server brute-force attack blocking whitelist, IPs with application attack permissions, and Web attack blocking whitelist.

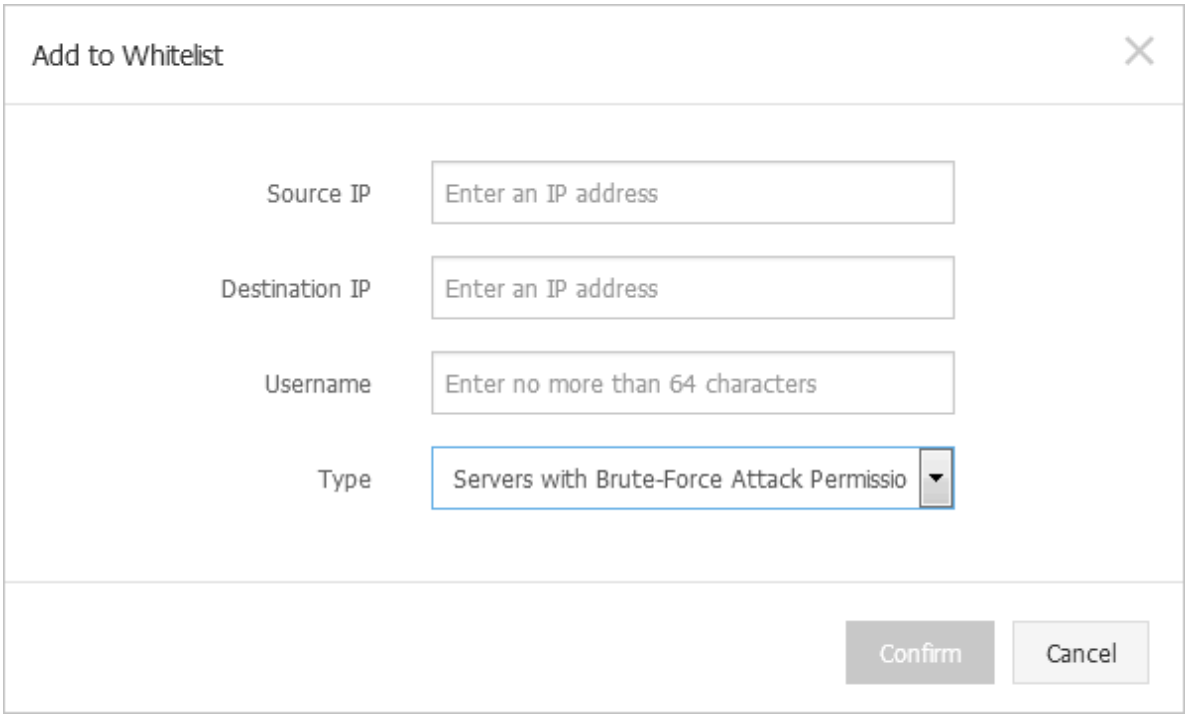
Context

If a normal request is regarded as an attack by the attack blocking function of Threat Detection Service or the unusual logon detection function of Server Guard, you can whitelist the source IP address to avoid further false positives.

Note:
Make sure that the whitelisted IP addresses can be trusted.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose System Management > Global Settings > Whitelist.**
3. **Click Add.**
4. **In the Add to Whitelist dialog box, set parameters for the whitelist entry.**



Parameter	Description
Source IP	The source IP address or CIDR block.

Parameter	Description
Destination IP	The destination IP address or CIDR block. **Note:** You can set this parameter when you configure `Servers with Brute-Force Attack Permissions` or `IPs with Application Attack Permissions`.
Username	The name of the user who adds the whitelist entry.
Type	• `Brute-Force Attack Blocking Whitelist`: Server Guard does not alert you on brute-force attacks or unusual logon events started by the IP addresses in this whitelist. • `Beaver WAF Whitelist`: The attack blocking function does not alert you on the Web attacks started by the IP addresses in this whitelist. • `Servers with Brute-Force Attack Permissions`: The attack blocking function does not alert you on brute-force attacks started by the IP addresses in this whitelist. • `IPs with Application Attack Permissions`: The traffic from the IP addresses in this whitelist is not detected as suspicious application attack traffic.

5. Click OK.

   You can click Delete to delete an unnecessary whitelist entry.

19.7.3.4 Physical machine protection

19.7.3.4.1 View and handle file tampering events

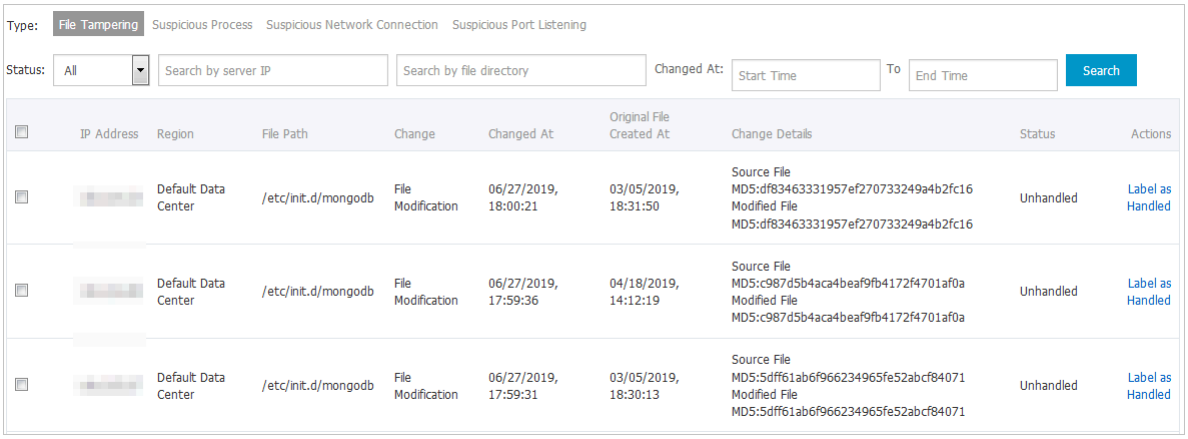
This topic describes how to check the integrity of files in specific directories of the server system, detect tampering events, and generate alerts.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose System Management > Global Settings > Physical Machine Protection, and click File Tampering.

3. **View file tampering events, as shown in** *Figure 19-14: File tampering events*.

   Figure 19-14: File tampering events



4. **Troubleshoot a file tampering event.**

   - **If you confirm that the event is an intrusion, take measures to enhance the server security, and analyze the cause of the intrusion.**
   - **If you confirm that the event is normal or is an intrusion that has been handled, click Label as Handled. In the message that appears, click Confirm to change the event status to Handled.**

19.7.3.4.2 View and handle suspicious processes

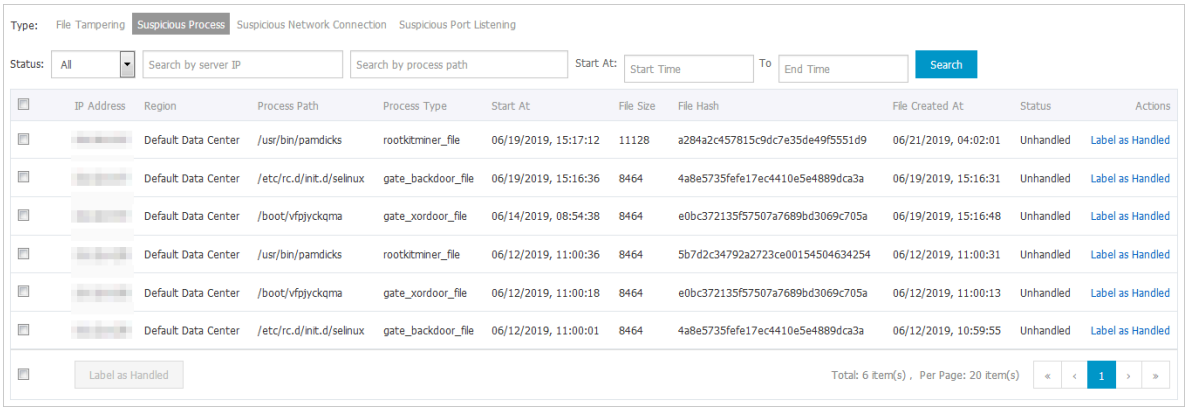
This topic describes how to detect suspicious processes and generate alerts.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose System Management > Global Settings > Physical Machine Protection, and click Suspicious Process.**

3. **View suspicious processes, as shown in** *Figure 19-15: Suspicious processes*.

   Figure 19-15: Suspicious processes



4. **Troubleshoot a suspicious process.**
   - **If you confirm that the process is suspicious, take measures to enhance the server security, and analyze the cause of the intrusion.**
   - **If you confirm that the process is normal or is a suspicious process that has been handled, click Label as Handled. In the message that appears, click Confirm to change the status to Handled.**

19.7.3.4.3 View and handle suspicious network connections

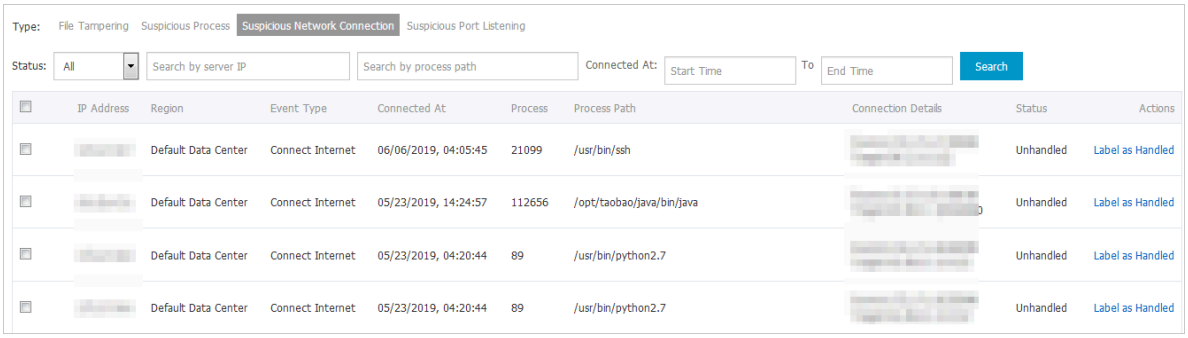
This topic describes how to detect active connections to public networks and generate alerts.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose System Management > Global Settings > Physical Machine Protection, and click Suspicious Network Connection.**

3. **View suspicious network connections, as shown in** *Figure 19-16: Suspicious network connections*.

   Figure 19-16: Suspicious network connections



4. **Troubleshoot a suspicious network connection.**
   - **If you confirm that the connection is suspicious, take measures to enhance server security, and analyze the cause of the intrusion.**
   - **If you confirm that the connection is normal or is a suspicious connection that has been handled, click Label as Handled. In the message that appears, click Confirm to change the event status to Handled.**

19.7.3.4.4 View and handle suspicious port listening events

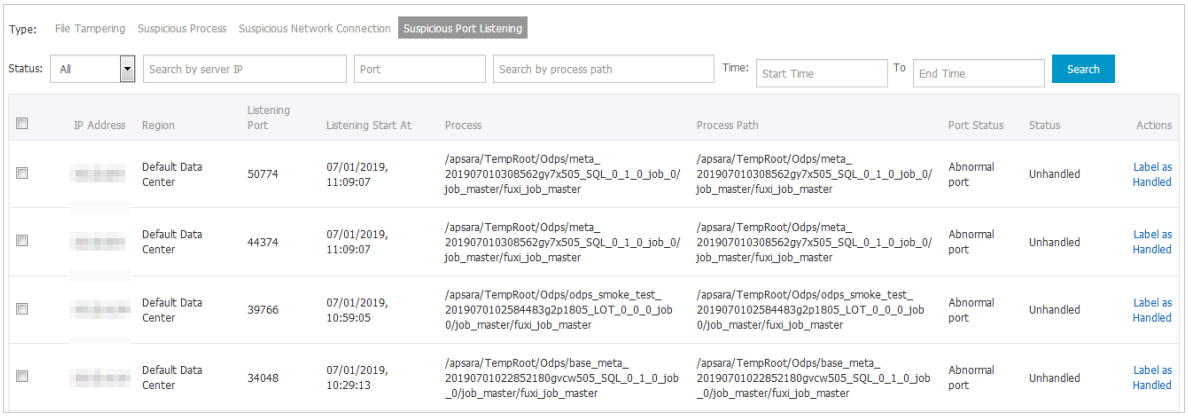
This topic describes how to detect suspicious port listening events and generate alerts accordingly.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose System Management > Global Settings > Physical Machine Protection, and click Suspicious Port Listening.**

3. **View suspicious port listening events, as shown in** *Figure 19-17: Suspicious port listening events*.

   Figure 19-17: Suspicious port listening events



4. **Troubleshoot a suspicious port listening event.**
   - **If you confirm that the port listening event is suspicious, take measures to enhance the server security, and analyze the cause of the intrusion.**
   - **If you confirm that the port listening event is normal or is a suspicious event that has been handled, click Label as Handled. In the message that appears, click Confirm to change the event status to Handled.**

19.8 Optional security products

19.8.1 Anti-DDoS settings

19.8.1.1 Overview

In Distributed Denial of Service (DDoS) attacks, attackers exploit the client-server model to combine multiple computers into a platform that can launch attacks on one or more targets. This greatly increases the threat of attacks.

Common DDoS attack types include:

- **Network-layer attacks: A typical example is UDP reflection attacks, such as NTP flood. These attacks use heavy traffic to congest the network of the victim, disabling proper responses to user requests.**
- **Transport-layer attacks: Typical examples include SYN flood and connection flood. These attacks consume a large number of connection resources of a server to cause denial of service.**

- Session-layer attacks: A typical example is SSL flood. These attacks consume the SSL session resources of a server to cause denial of service.
- Application-layer attacks: Typical attack types include DNS flood, HTTP flood, and game zombie attacks. These attacks consume a large amount of application processing resources of a server to cause denial of service.

Apsara Stack Security can redirect, scrub, and re-inject attack traffic to protect your server against DDoS attacks and ensure normal business operations.

Note:
Apsara Stack Security cannot scrub the traffic between internal networks.

19.8.1.2 View DDoS events

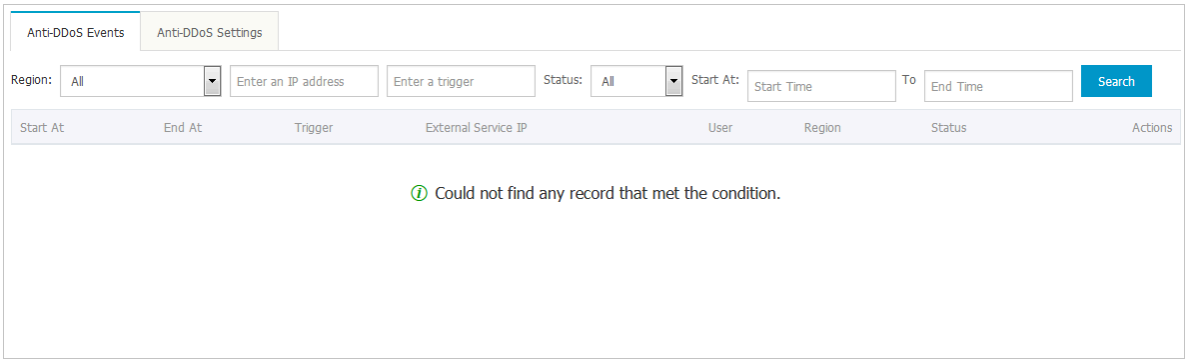
This topic describes how to view distributed denial of service (DDoS) events.

Context

During or after the traffic scrubbing process, Apsara Stack Security reports security events to Apsara Stack Security Center.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose Network Security > Anti-DDoS Service. Then, click Anti-DDoS Events.



3. Set the search criteria and click Search.

   The system returns a list of DDoS events that meet the search criteria.

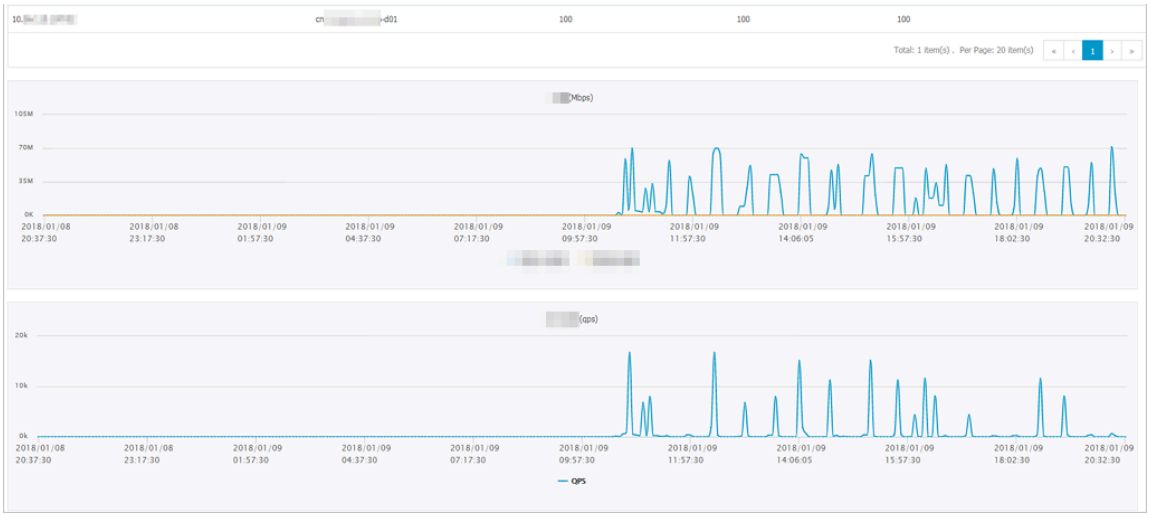
Parameter	Description
Trigger	The metric that exceeds the configured alert threshold in the DDoS attack traffic.
External Service IP	The IP address that was under a DDoS attack.

Parameter	Description
Status	· Scrubbing: indicates that traffic scrubbing is in progress . · Scrubbed: indicates that traffic scrubbing is complete.
Actions	· Cancel Scrubbing: allows you to stop traffic scrubbing. · Traffic Analysis: allows you to view the traffic protocol and top 10 attacking servers in the DDoS event. · View Traffic: allows you to view the alert thresholds and traffic diagram.

4. **View and analyze a DDoS event.**

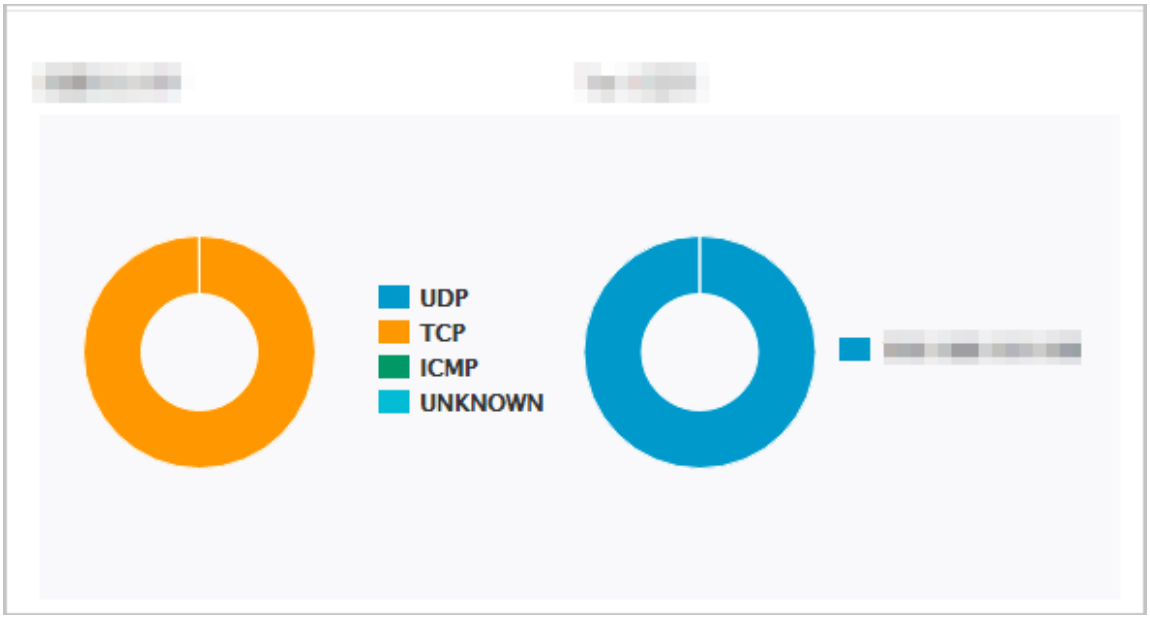
   - **Click View Traffic to view the alert thresholds and traffic diagram of the corresponding IP address.**

     Figure 19-18: Traffic diagram



   - **Click Traffic Analysis to view the traffic protocol and top 10 attacking servers in the DDoS event.**

     Figure 19-19: Traffic Analysis page



19.8.1.3 Anti-DDoS rules

19.8.1.3.1 Add an anti-DDoS rule

This topic describes how to add an anti-DDoS rule to set DDoS alert thresholds for an IP address or a Classless Inter-Domain Routing (CIDR) block. After the alert thresholds are set, the system detects DDoS attacks to the IP address or CIDR block based on the specified alert thresholds. If no alert thresholds are set for an IP address or CIDR block, the system detects DDoS attacks to the IP address or CIDR block based on the global alert thresholds.

Context

After an alert threshold of DDoS traffic is set for an IP address, an alert is triggered when the traffic to the IP address reaches the threshold. The alert thresholds for an IP address must be set based on the traffic volume. Excessive traffic volume indicates a possible DDoS attack. We recommend that you set an alert threshold to a value slightly higher than the peak traffic volume.

Apsara Stack Security supports global alert thresholds or alert thresholds for a specific CIDR block or IP address.

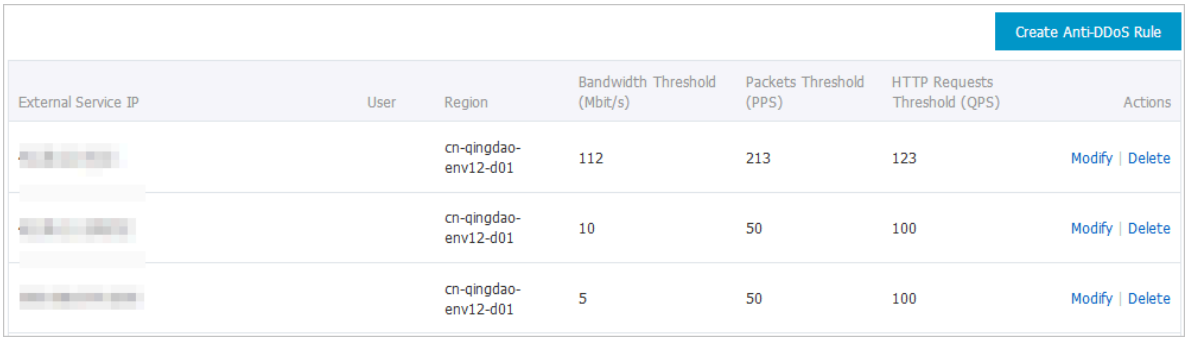
- **Global alert threshold: You cannot add a global alert threshold. It is set during service initialization.**
- **Alert threshold for a specific CIDR block: You can set an alert threshold for a specific CIDR block based on the traffic volume. Compared with global alert thresholds, CIDR block-specific thresholds allow you to precisely control the traffic to each CIDR block.**
- **Alert threshold for a specific IP address: You can set an alert threshold for a specific IP address based on the traffic volume. Compared with CIDR block-specific thresholds, IP address-specific thresholds allow you to precisely control the traffic to each IP address.**

Procedure

1. *Log on to Apsara Stack Security Center*.

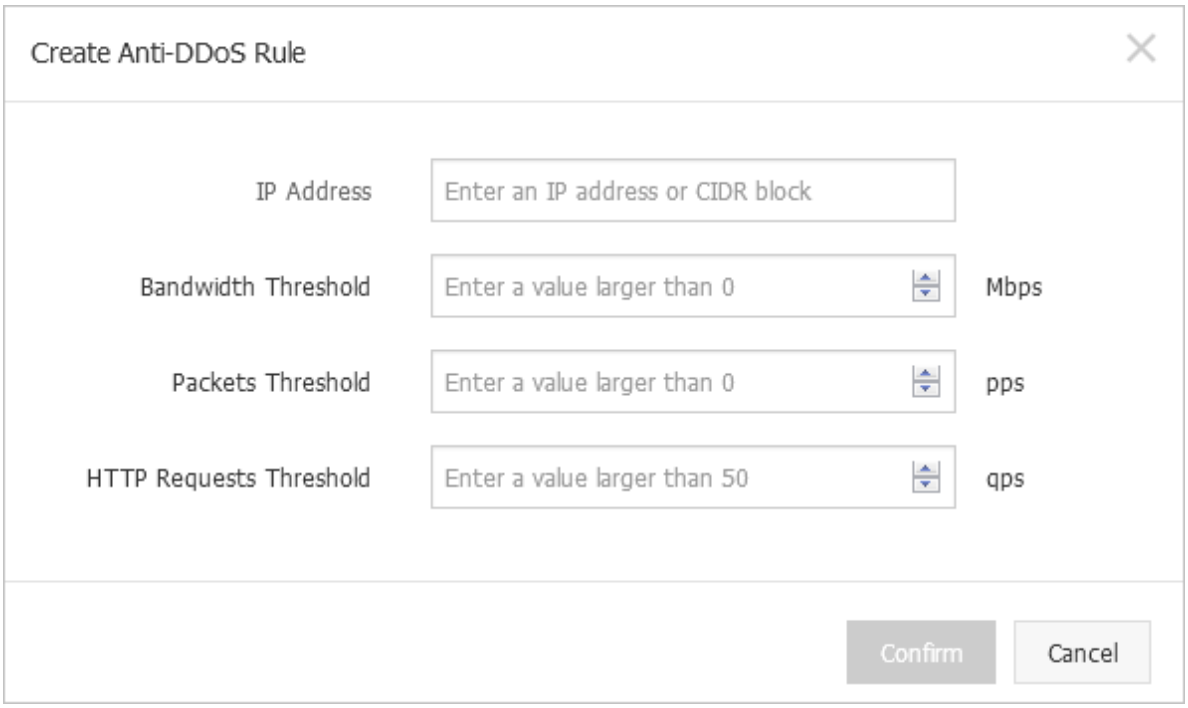
2. **Choose Network Security > Anti-DDoS Service. Then, click Anti-DDoS Settings.**

   Figure 19-20: Threshold list



3. **Click Create Anti-DDoS Rule.**
4. **In the Create Anti-DDoS Rule dialog box that appears, set the rule parameters.**

   Figure 19-21: Create Anti-DDoS Rule dialog box



Parameter	Description
IP	The IP address or CIDR block to which the alert thresholds are applied. **Note:** Before setting the alert thresholds, make sure that the corresponding CIDR block has been added in *Add a CIDR block for traffic monitoring*.

Parameter	Description
Bandwidth Threshold	The alert threshold for bandwidth usage in a data center. When the inbound or outbound traffic rate reaches this threshold, DDoS detection is triggered. Generally, set this parameter to a value slightly higher than the traffic peak. We recommend that you set this parameter to 100 or higher. Unit: Mbit/s.
Packets Threshold	The alert threshold for the packet transmission rate in a data center. When the inbound or outbound packet transmission rate reaches this threshold, DDoS detection is triggered. Generally, set this parameter to a value slightly higher than the traffic peak. We recommend that you set this parameter to 20000 or higher. Unit: packets per second (PPS).
HTTP Requests Threshold	The alert threshold for the rate at which the servers in a data center receive HTTP requests. When the inbound and outbound HTTP request rate reaches this threshold, DDoS detection is triggered. Generally, set this parameter to a value slightly higher than the traffic peak. We recommend that you set this parameter to 100000 or higher. Unit: queries per second (QPS).

5. Click OK.

19.8.1.3.2 Manage anti-DDoS rules

This topic describes how to modify or delete anti-DDoS rules.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose Network Security > Anti-DDoS Service. Then, click Anti-DDoS Settings.

3. Manage an anti-DDoS rule by clicking a button in the Actions column.

   - Modify the anti-DDoS rule.

     Click Modify. In the Modify Anti-DDoS Rule dialog box that appears, modify the alert thresholds and click OK.

   - Delete the anti-DDoS rule.

     Click Delete to delete the anti-DDoS rule.

     > **Note:**
     > The default anti-DDoS rules cannot be deleted.

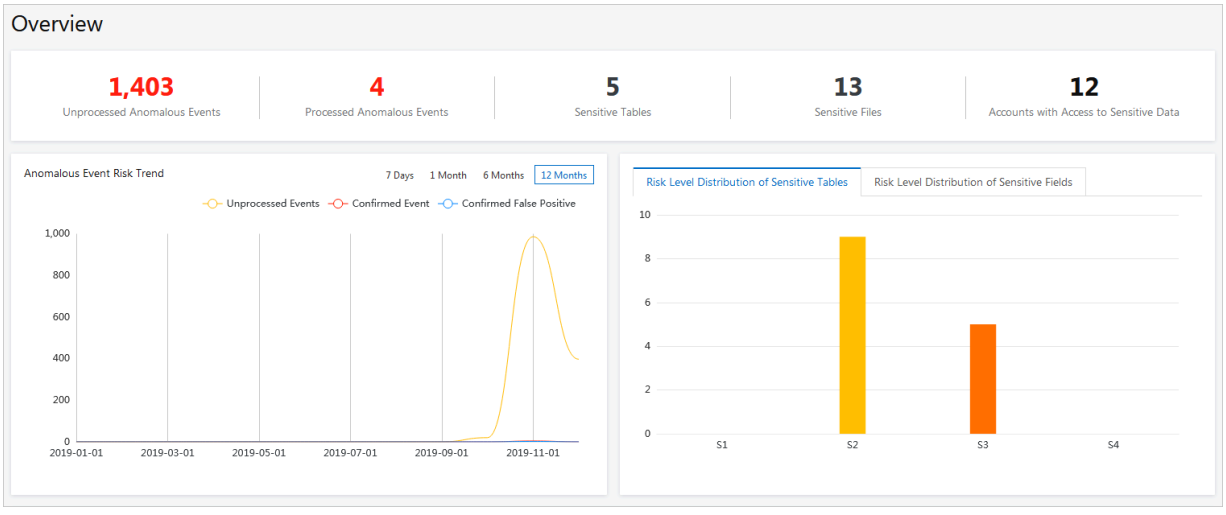
19.8.2 Sensitive Data Discovery and Protection

19.8.2.1 Overview

This topic describes the overview page of Sensitive Data Discovery and Protection (SDDP). This page displays the overall security status of data protected by SDDP, and allows a security administrator to quickly understand the current security status of sensitive data.

SDDP can detect sensitive data in your data assets based on sensitive data detection rules and track the use of sensitive data. SDDP also provides a data overview for you to obtain the security status of your data assets in real time.

Choose Data Security > Sensitive Data Discovery and Protection > Overview. On the Overview page, view the overall security status of the sensitive data.



- Overview: displays the overall information of sensitive data, including the number of unprocessed anomalous activities, the number of anomalous

activities confirmed as violations, the total number of sensitive tables, the total number of sensitive objects, and accounts that accessed sensitive data.

- Abnormal event risk trend: displays the trends of anomalous activities in a line chart. You can select `7 days`, `1 month`, `6 months`, or `12 months` to view the trends of unprocessed anomalous activities, anomalous activities confirmed violations, and anomalous activities excluded as false positives.
- Sensitive table risk level distribution: displays the distribution of sensitive tables at the S1, S2, S3, and S4 risk levels.
- Sensitive field risk level distribution: displays the distribution of sensitive fields at the S1, S2, S3, and S4 risk levels.
- Data flow situation:
  - Displays the dynamic statistics on core data flows in Datahub and Cloud Data Pipeline (CDP).
  - Provides a data flow chart that dynamically displays the data flow status and abnormal output. You can click an anomalous activity in the flow chart to go to the Abnormal data flow page.

    Monitors the data links among entities such as data storage services MaxCompute, , Object Storage Service (OSS), and Table Store, data transmissi on services Datahub and CDP, the data flow processing service Blink, external databases, and external files.

19.8.2.2 Process anomalous activities

This topic describes how to process anomalous activities in Sensitive Data Discovery and Protection (SDDP). SDDP can detect anomalous activities related to sensitive data and generate alerts. On the Exception event handling page, you can confirm anomalous activities as violations or exclude anomalous activities as false positives.

Context

SDDP divides anomalous activities into the following types:

- Anomalous permission access: Permissions are used anomalously. For example, a user logs on from an unusual IP address or by using the AccessKey of another user.

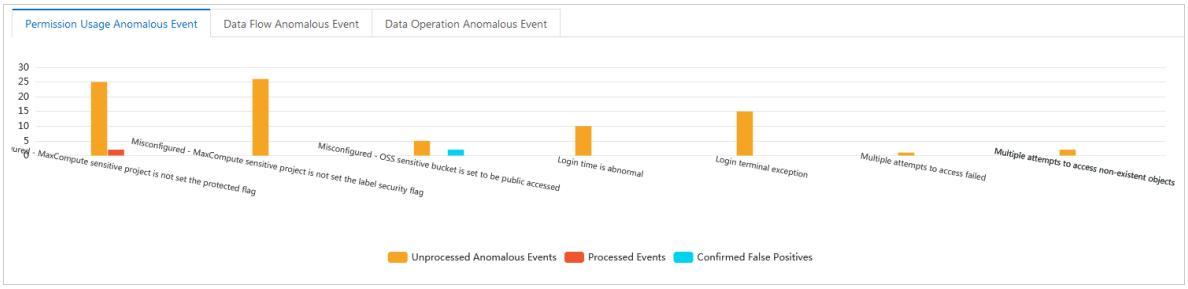
- **Anomalous data flow: Anomalous activities are detected during data flows. For example, a user downloads sensitive data files unnecessarily or during an unusual time period.**
- **Anomalous data operation: Anomalous operations are performed on sensitive data. For example, a user modifies sensitive fields.**

Procedure

1. *Log on to Apsara Stack Security Center*.
2. **Choose Data Security > Sensitive Data Discovery and Protection > Exception event handling.**
3. **View the statistics on anomalous activities.**

   You can view the statistics on different types of anomalous activities in the upper part of the Exception event handling page. The statistics include the anomalous activity types, number of processed anomalous activities, and number of unprocessed anomalous activities.

   Figure 19-22: Statistics on anomalous activities

   a) **Click the Abnormal permissions use, Abnormal data flow, or Abnormal data operation tab.**

   b) **View the bar chart of the statistical results.**

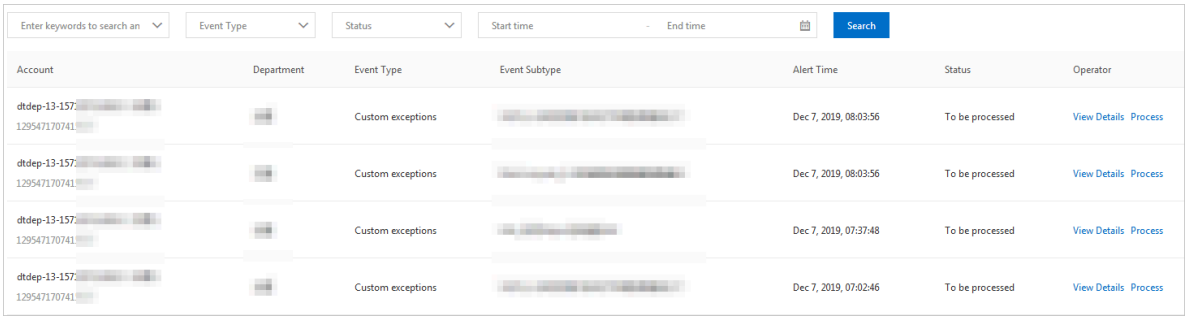
   c) **Move the pointer over an anomalous activity to view the detailed information.**

4. **Set the search criteria and click Search to search for anomalous activities.**

   You can filter anomalous activities by keyword, department, type, status, and alerting time.

5. Process anomalous activities.

   You can process anomalous activities detected by SDDP in the anomalous activity list in the lower part of the Exception event handling page.



   a) Find the target anomalous activity and click View details in the Actions column to view the details of the anomalous activity.

   b) Find the target exception and click Do in the Actions column to process the anomalous activity.

   c) In the Abnormal event handling dialog box that appears, process the anomalous activity.

Parameter	Description
Add Processing Record	Check the anomalous activity and record the verification process.
Verification Result	· Confirmed and Processed: Confirm the anomalous activity as a violation. If you select this option without manually processing the anomalous activity in the corresponding service, SDDP keeps generating alerts for the anomalous activity. · False Positive: Exclude the anomalous activity as a false positive. After you select this option, SDDP no longer generates alerts for this anomalous activity. That is, this anomalous activity will no longer appear on the Exception event handling page.

Parameter	Description
Anomalous Event Sample-based Enhancement	Specify whether to return the processing result of the anomalous activity as a feedback to the detection algorithm. **Note:** If you select this check box: · An anomalous activity that is excluded as a false positive will be returned as a feedback to the current algorithm as a false positive sample. · An anomalous activity that is confirmed as a violation will be returned to the current algorithm as a positive sample. This improves the accuracy of anomalous activity detection, but may also increase the missing rate.

d) Click Complete.

19.8.2.3 Detect sensitive data

19.8.2.3.1 Sensitive data overview

This topic describes how to view the overall security status of your data assets.

Choose Data Security > Sensitive Data Discovery and Protection > Sensitive data identification > Sensitive data overview. On the Sensitive data overview page, you can view the overall security status of your data assets.

- You can view the overall information about sensitive data. The information includes the total numbers of tables, objects, sensitive instances, sensitive tables , and sensitive objects.
- You can search for sensitive data based on conditions such as the risk level, asset type, sensitive data type, and asset name.
- You can view the statistics on the authorization information and sensitive data in Apsara Stack services such as MaxCompute, Object Storage Service (OSS), and Table Store in real time.

19.8.2.3.2 View the statistics on sensitive data of MaxCompute

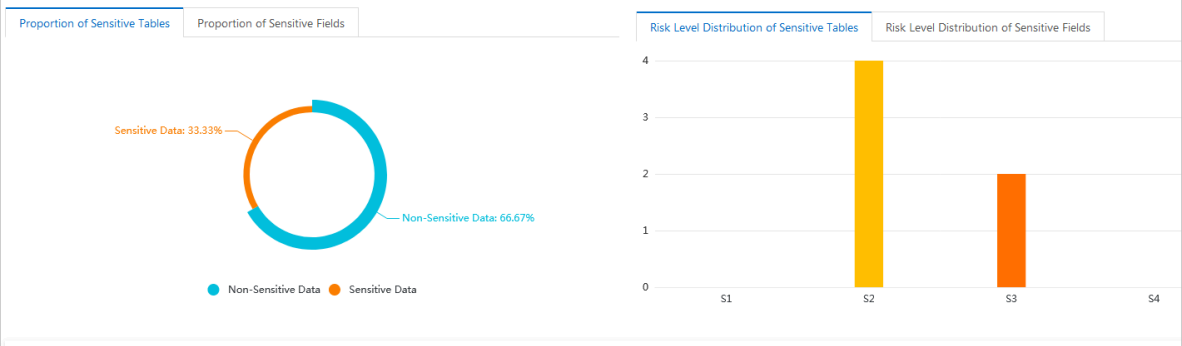
This topic describes how to view statistics on sensitive data of MaxCompute.

Context

MaxCompute is a rapid and fully-managed data warehouse solution that can process terabytes or petabytes of data. MaxCompute provides you with complete data import schemes and various classic distributed computing models. It supports fast computing on a large amount of data, effectively saves costs for enterprises, and guarantees data security.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. **Choose Data Security > Sensitive Data Discovery and Protection > Sensitive data identification > MaxCompute.**
3. **View the statistics on sensitive data of MaxCompute.**



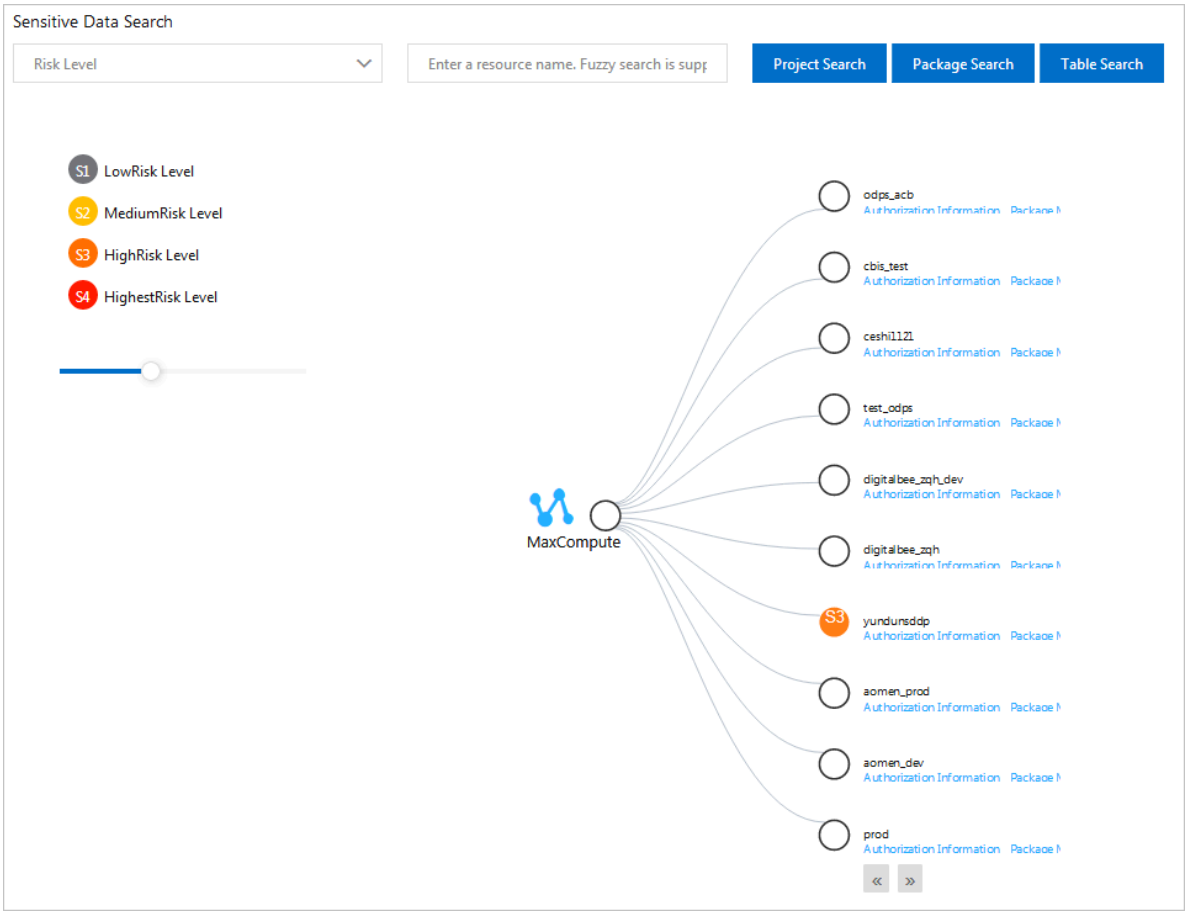
a) **In the Sensitive data statistics section, enter a keyword and select the target MaxCompute project from the drop-down list.**

> **Note:**
> **To view the statistics on all MaxCompute projects, select `All` from the drop-down list.**

b) **On the Sensitive table ratio and Sensitive field ratio tabs, view the percentages of sensitive and non-sensitive tables and fields.**

c) **On the Sensitive table risk level distribution and Sensitive field risk level distribution tabs, view the distribution of sensitive tables and fields at the S1, S2, S3, and S4 risk levels.**

4. Query the sensitive data of MaxCompute.



By default, the system displays all MaxCompute projects. The system displays different risk levels in different colors. To view the information about a specific project, package, table, or field, follow these steps:

a) Select a risk level from the Risk level drop-down list.

b) Enter a keyword of the project, package, or table in the search field.

c) Click Search project, Search package, or Search table.

You can view the relationships among the projects, packages, tables, and fields, and the related authorization information in a tree map.

- The tree map displays the distribution of sensitive data in MaxCompute.
- You can view the authorization information of a project, package, table, or field. The system displays the authorization information by category, including the authorized users and violations.
- You can click Package management under a project to view the packages in the project, including the tables and fields in the packages and related authorization information.

d) Move the pointer over the project, package, table, or field to view its details.

19.8.2.3.3 View the statistics on sensitive data of Table Store

This topic describes how to view the statistics on sensitive data of Table Store.

Context

Table Store is a multi-model NoSQL database service developed by Apsara Stack. Table Store can store a large amount of structured data and support fast query and analysis. The distributed storage and powerful index-based search engine enable Table Store to store petabytes of data while guaranteeing a 10 million TPS and a latency within milliseconds.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. **Choose Data Security > Sensitive Data Discovery and Protection > Sensitive data identification > OTS.**
3. **View the statistics on sensitive data of Table Store.**

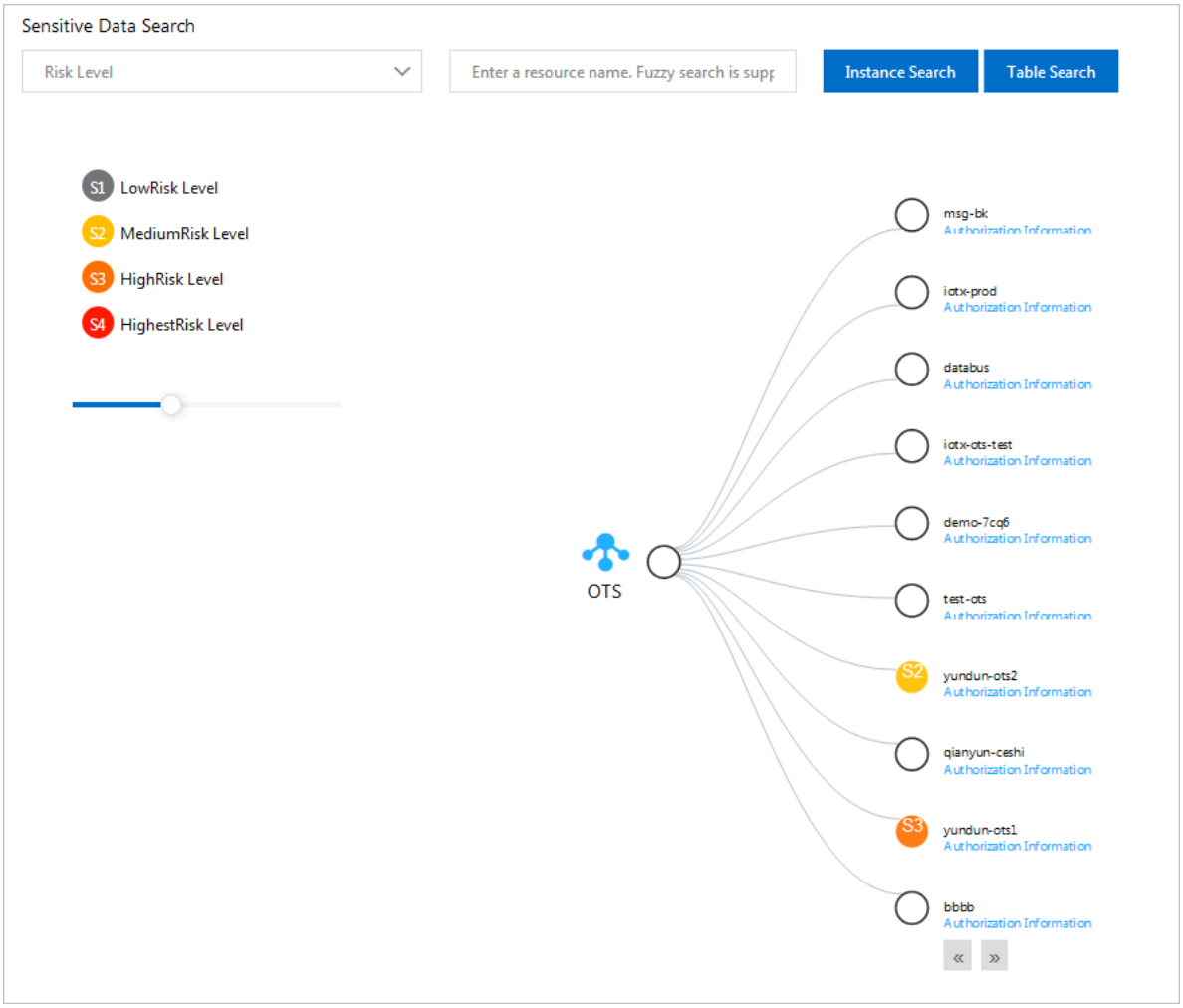
a) **In the Sensitive data statistics section, enter a keyword and select the target Table Store instance from the drop-down list.**

> **Note:**
> **To view the statistics on all Table Store instances, select `All` from the drop-down list.**

b) **On the Sensitive table ratio and Sensitive field ratio tabs, view the percentages of sensitive and non-sensitive tables and fields.**

c) **On the Sensitive table risk level distribution and Sensitive field risk level distribution tabs, view the distribution of sensitive tables and fields at the S1, S2, S3, and S4 risk levels.**

4. Query the sensitive data of Table Store.



By default, the system displays all Table Store instances. To view the information about a specific instance or table, follow these steps:

a) Select a risk level from the Risk level drop-down list.

b) Enter a keyword of the instance or table in the search field.

c) Click Search instances or Search Tables.

- The tree map displays the distribution of sensitive data in Table Store.
- You can view the authorization information of an instance or a table. The system displays the authorization information by category, including the authorized users and violations.

d) Move the pointer over the instance or table to view its details.

19.8.2.3.4 View the statistics on sensitive data of OSS

This topic describes how to view the statistics on sensitive data of Object Storage Service (OSS).

Context

OSS is a secure and reliable cloud storage service provided by Apsara Stack. It can store a large amount of data at low costs. OSS can store any type of file and is therefore suitable for various websites, enterprises, and developers.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Data Security > Sensitive Data Discovery and Protection > Sensitive data identification > OSS.**
3. **View the statistics on sensitive data of OSS.**

View the charts on the Sensitive file object ratio and Sensitive object risk level distribution tabs.

To view the charts for a specific OSS bucket, follow these steps:

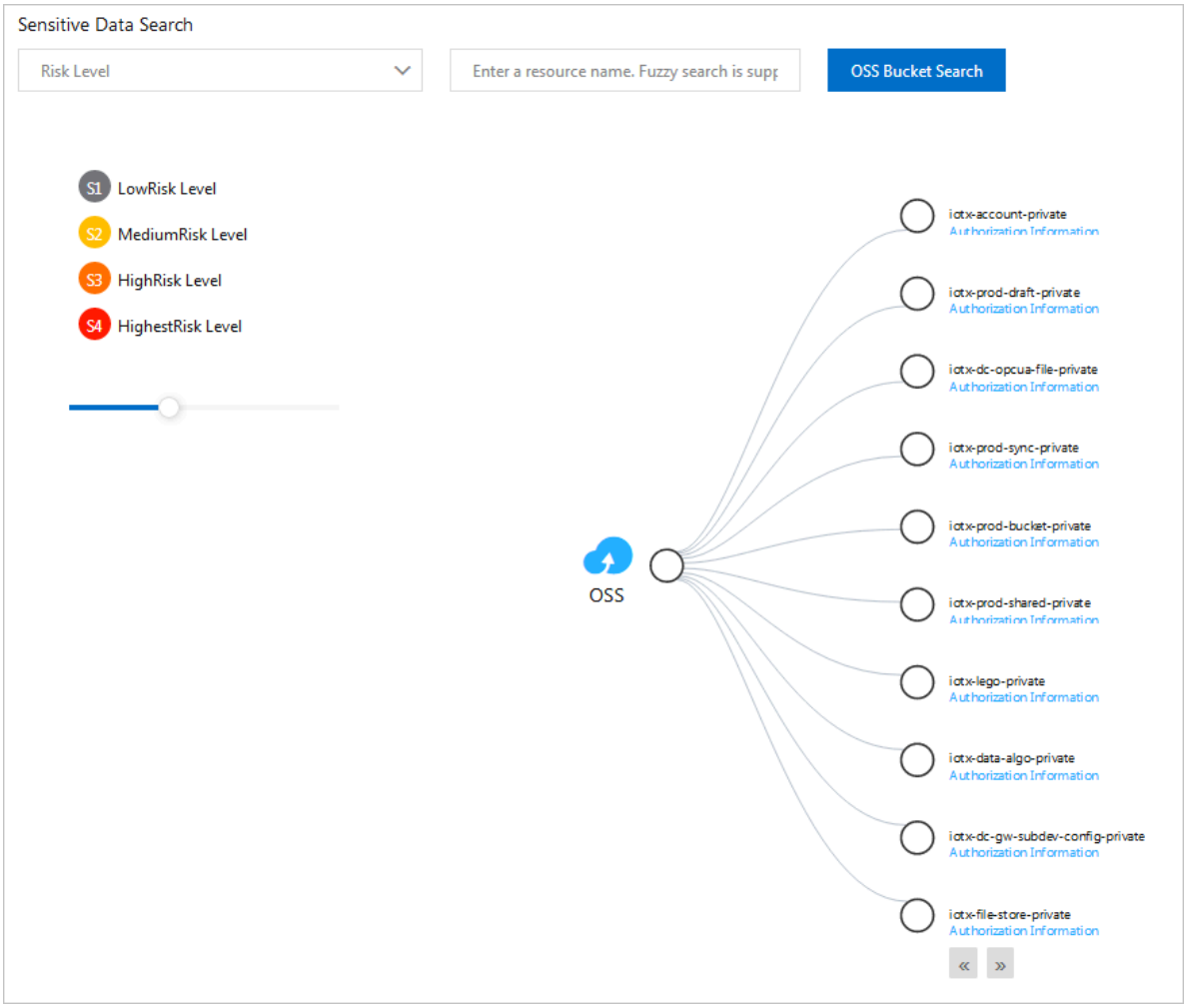
a) **In the Sensitive data statistics section, enter a keyword and select the target OSS bucket from the drop-down list.**

> **Note:**
> **To view the statistics on all OSS buckets, select `All` from the drop-down list.**

b) **On the Sensitive file object ratio tab, view the percentages of sensitive and non-sensitive objects.**

c) **On the Sensitive object risk level distribution tab, view the distribution of sensitive objects at the S1, S2, S3, and S4 risk levels.**

4. Query the sensitive data of OSS.



By default, the system displays all OSS buckets. To view the information about a specific bucket, follow these steps:

a) Select a risk level from the Risk level drop-down list.

b) Enter a keyword of the bucket in the search field and click Search bucket.

- The tree map displays the distribution of sensitive data in OSS.
- You can view the authorization information of a bucket. The system displays the authorization information by category, including the authorized users and violations.

c) Move the pointer over the bucket to view its details.

19.8.2.4 Check data permissions

19.8.2.4.1 View permission statistics

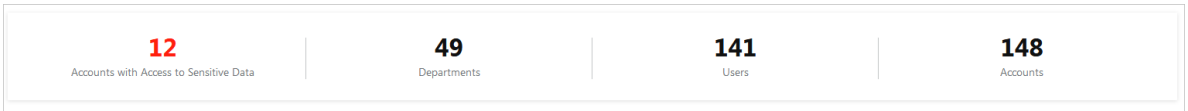
This topic describes how to view permission statistics.

Context

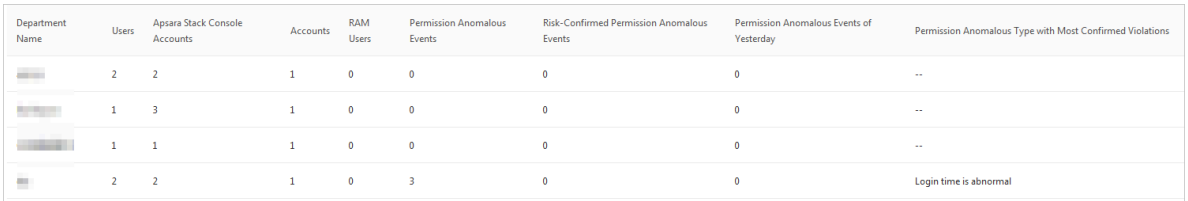
On the Permissions management page, you can check the overall permission distribution of Apsara Stack. With this feature, you can quickly identify high-risk accounts and users, and troubleshoot and resolve security issues in a timely manner.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Data Security > Sensitive Data Discovery and Protection > data permission > Permissions management.
3. View the overall statistics on permissions.



   - Number of accounts accessible to sensitive data: the number of accounts that can access sensitive data.
   - Total number of departments: the number of departments in Apsara Stack.
   - Total number of people: the number of users in Apsara Stack.
   - Total number of accounts: the number of accounts in Apsara Stack.
4. View the department-level statistics on permissions.



Department Name	Users	Apsara Stack Console Accounts	Accounts	RAM Users	Permission Anomalous Events	Risk-Confirmed Permission Anomalous Events	Permission Anomalous Events of Yesterday	Permission Anomalous Type with Most Confirmed Violations
[redacted]	2	2	1	0	0	0	0	--
[redacted]	1	3	1	0	0	0	0	--
[redacted]	1	1	1	0	0	0	0	--
[redacted]	2	2	1	0	3	0	0	Login time is abnormal

You can view the statistics on the users, accounts, and anomalous activities related to permission access for each department.

19.8.2.4.2 Query permissions

This topic describes how to query permissions.

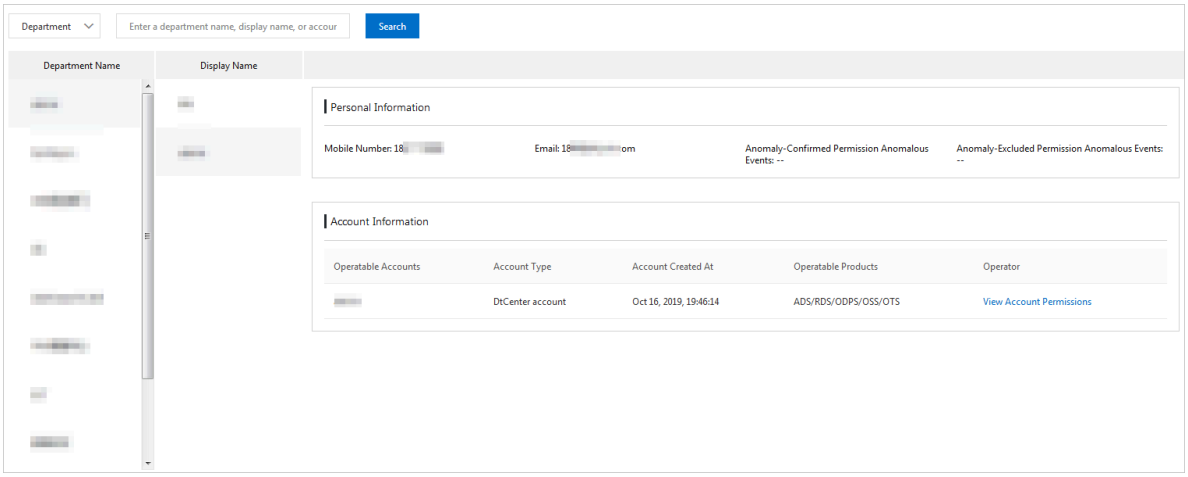
Context

You can search for an account to view the account information. With this feature, you can quickly find the owner for sensitive data.

Procedure

1. *Log on to Apsara Stack Security Center.*

2. Choose Data Security > Sensitive Data Discovery and Protection > data permission > Permissions search.



3. Search for the target account.

   To search for an account, follow these steps:

   a) Select a department or user from the drop-down list.

   b) Enter a keyword in the search field.

   c) Click Search. The accounts containing the keyword are listed in the Display name column.

   You can also click a department in the Department name column. All accounts of the department are listed in the Display name column.

4. In the Display name column, click the target account.

5. In the right pane, view the information in the Personal information and Accounts sections.

   - Personal information

     You can view the contact information of the account owner, the number of confirmed anomalous activities related to permission access, and the number of excluded anomalous activities related to permission access.

   - Accounts

     You can view the accounts that the owner can use, the type and creation time of the accounts, and Apsara Stack services that the accounts can access.

     You can click View account permissions in the Actions column for an account to view the resources, resource types, resource paths, and operation permissions of the account.

19.8.2.5 Monitor data flows

19.8.2.5.1 View data flows in Datahub

This topic describes how to view data flows in Datahub.

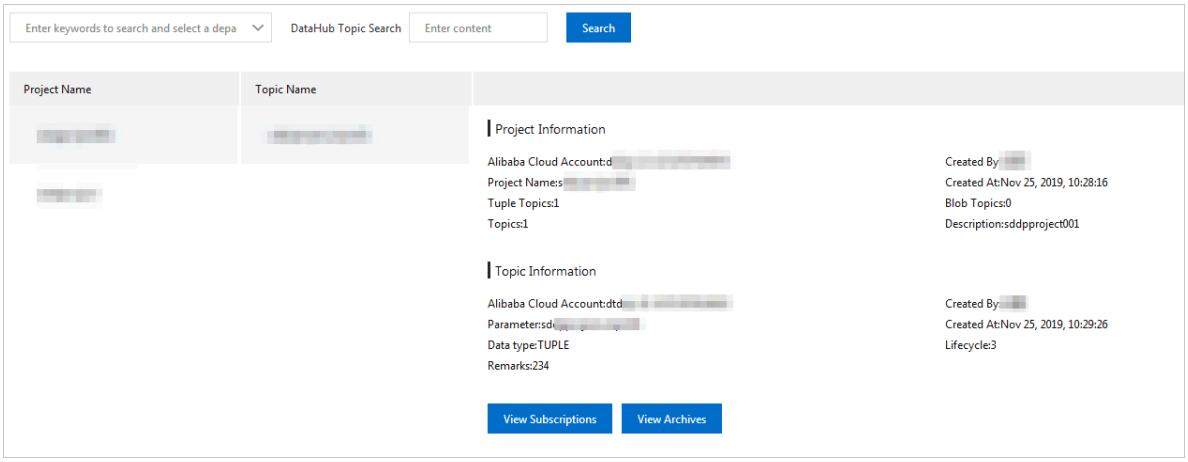
Context

Datahub is a platform designed to process streaming data. You can publish and subscribe to applications for streaming data in Datahub and distribute the data to other platforms. Datahub allows you to analyze streaming data and build applicatio ns based on the streaming data.

On the DataHub page, you can view the details of data flows in Datahub, including the relationships between Datahub projects and topics, and the relationships among topics, subscribed applications, and archive sources.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Data Security > Sensitive Data Discovery and Protection > Data flow monitoring > DataHub.**
3. **Enter a keyword and select a department from the drop-down list, enter a keyword in the DataHub topic search field, and then click Search.**



Note:

You can also click the target project in the project name column and then click the target topic in the topic name column.

You can view the information about the topic in the Project information and Topic information sections.

- Project information

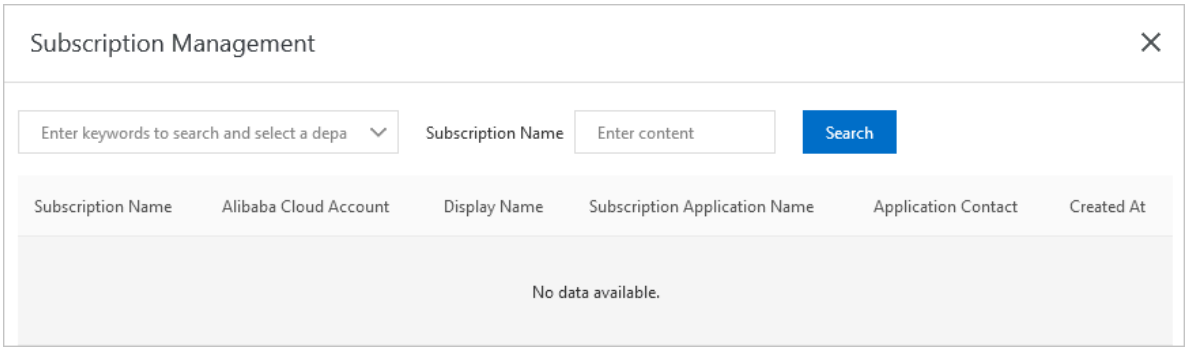
  Displays the information such as the project name, Apsara Stack account, creator, creation date, and number of topics.

- Topic information

  Displays the information such as the project name, Apsara Stack account, creator, creation date, and type.

4. Click View subscription to view the subscription list.

   The subscription list contains the information such as the subscription name, Apsara Stack account of the creator, display name, name of the subscribed application, and application contact.



   a) Enter a keyword and select a department from the drop-down list.
   b) Enter a keyword in the DataHub topic search field.
   c) Click Search to find the target Datahub topic.

5. Click View archive to view the archive list.

   The archive list contains the information such as the name of the connected instance, Apsara Stack account of the creator, display name, source service, resource path, and risk level.

   a) Enter a keyword and select a department from the drop-down list.
   b) Enter a keyword in the DataHub topic search field.
   c) Click Search to find the target Datahub topic.

19.8.2.5.2 View data flows in CDP

This topic describes how to view data flows in Cloud Data Pipeline (CDP).

Context

DataWorks is a smart cloud R&D platform that provides big data operating system capabilities and professional, efficient, secure, and reliable services in an all-in-one manner. It can meet your requirements for data governance and quality management, allowing you to provide data services for external systems.

On the CDP flow monitoring page, you can view the details of data flows in CDP. The information includes the instance name, start time and end time of the instance, node ID, source location, destination location, and amount of synchronized data. In addition, you can click the instance name to go to the node details page. On this page, you can view the node execution link and the data mapping between the source and destination tables in real time.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Data Security > Sensitive Data Discovery and Protection > Data flow monitoring > CDP flow monitoring.
3. In the CDP flow monitoring section, select Today, This Week, or This Month and view information in Sync Instances: Source and Sync Instances: Target.
4. In the Sync Instances section, view the information such as the instance name, start time and end time of the instance, and node ID.
5. Click the instance name to go to the node details page. On this page, you can view the node execution link and the data mapping between the source and destination tables in real time.

19.8.2.6 Manage rules

19.8.2.6.1 Manage sensitive data detection rules

This topic describes how to manage sensitive data detection rules.

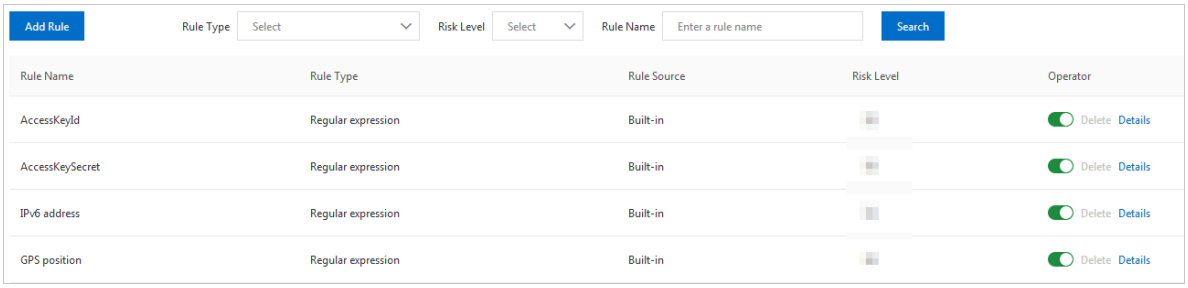
Context

Sensitive Data Discovery and Protection (SDDP) can detect sensitive data in Apsara Stack services such as MaxCompute, Object Storage Service (OSS), and Table Store.

SDDP detects sensitive data based on sensitive data detection rules. You can use the built-in rules provided by SDDP or configure custom rules based on your business needs to detect specific sensitive data.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Data Security > Sensitive Data Discovery and Protection > Rule configuration.
3. Click the Sensitive definition rule tab to view the existing sensitive data detection rules.



You can set Rule type, Risk level, and Rule name, and then click Search to search for rules.

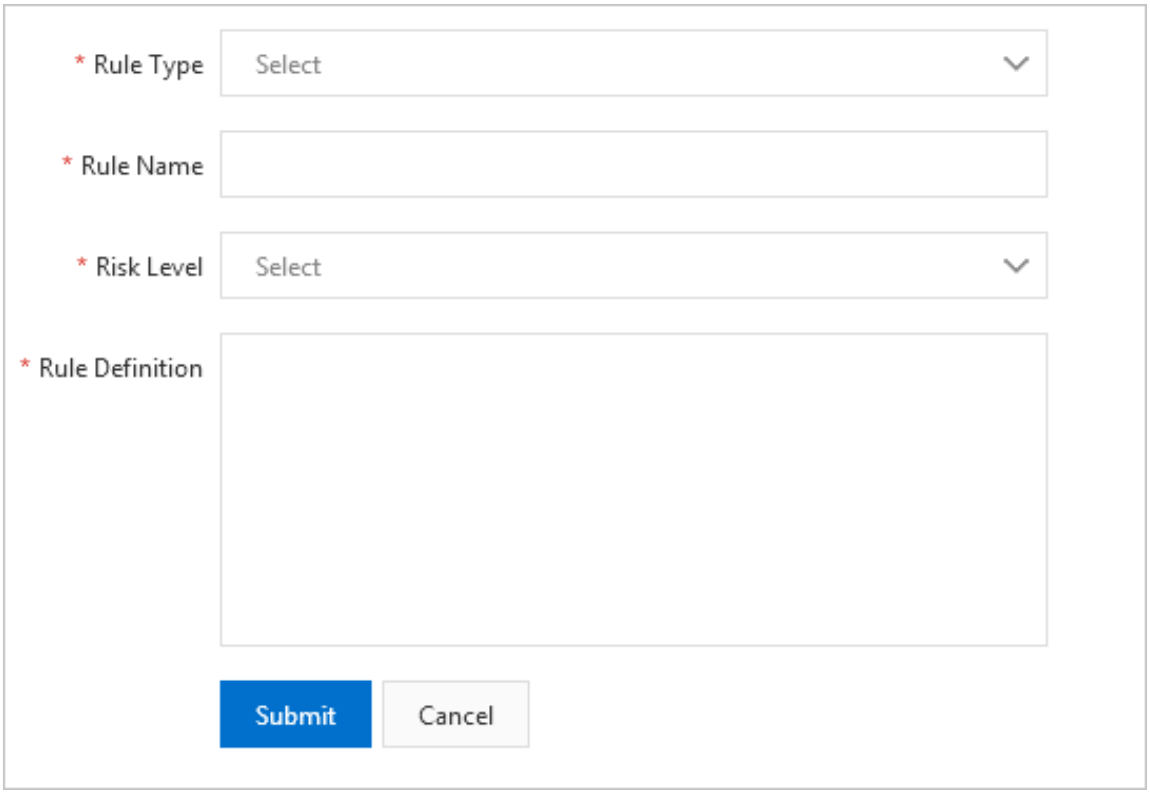
SDDP provides built-in algorithms for discovering sensitive data, and can use file clustering, deep neural network, and machine learning to detect sensitive images, texts, and fields. The built-in algorithms can detect sensitive data such as ID card numbers, addresses, phone numbers, and bank card numbers.

4. **Create a sensitive data detection rule.**

   You can create a custom rule to detect specific sensitive data.

   a) Click Add new rule.

   b) In the Add new rule dialog box, set the rule parameters.



   - **Rule type: the type of the custom rule. Valid values:** `Keyword` **and** `Regular expression`**.**
   - **Rule name: the name of the custom rule. We recommend that you name the rule based on its purpose.**
   - **Risk level: the risk level of the custom rule. Valid values: S1 (low risk), S2 (medium risk), S3 (high risk), and S4 (highest risk).**
   - **Rule definition: the content of the custom rule.**

   c) Click Submit.

5. **Manage sensitive data detection rules.**

   In the rule list, you can disable, enable, or delete a rule by clicking the corresponding button in the Actions column.

   Note:

- You can delete sensitive data detection rules, but cannot modify them. After a sensitive data detection rule is deleted, SDDP no longer detects corresponding data as sensitive data. Exercise caution when deleting a sensitive data detection rule.
- A sensitive data detection rule is enabled by default after it is created. If you do not regard certain data as sensitive data, you can disable the corresponding sensitive data detection rule. After a sensitive data detection rule is disabled, SDDP no longer detects corresponding data as sensitive data. We recommend that you enable all sensitive data detection rules to reduce risks.
- Rules for which `Built-in` appears in the Cloud account column are default rules. If no custom rules are configured, SDDP can still detect sensitive data based on these default rules. You cannot modify or delete the default rules.

19.8.2.6.2 Manage sensitive data definition rules

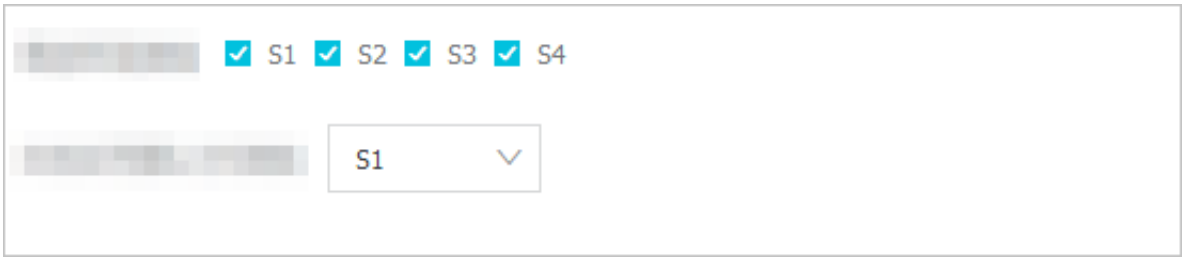
This topic describes how to manage sensitive data definition rules.

Context

You can use sensitive data definition rules to specify risk levels for sensitive data.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Data Security > Sensitive Data Discovery and Protection > Rule configuration.
3. Click the Sensitive definition rule tab to view the risk levels.



4. In the Please select a security level section, select the risk levels of sensitive to be used in Sensitive Data Discovery and Protection (SDDP).
5. In the The default marking level if not effectively recognized drop-down list, select any level from `S1` to `S4`.

Note:

The specified level indicates the default risk level of data that has not been identified as sensitive data.

19.8.2.6.3 Manage the thresholds and rules for detecting anomalous activities

This topic describes how to manage the thresholds and rules for detecting anomalous activities.

Context

On the Abnormal output configuration page, you can customize the thresholds and rules for detecting anomalous activities.

Procedure

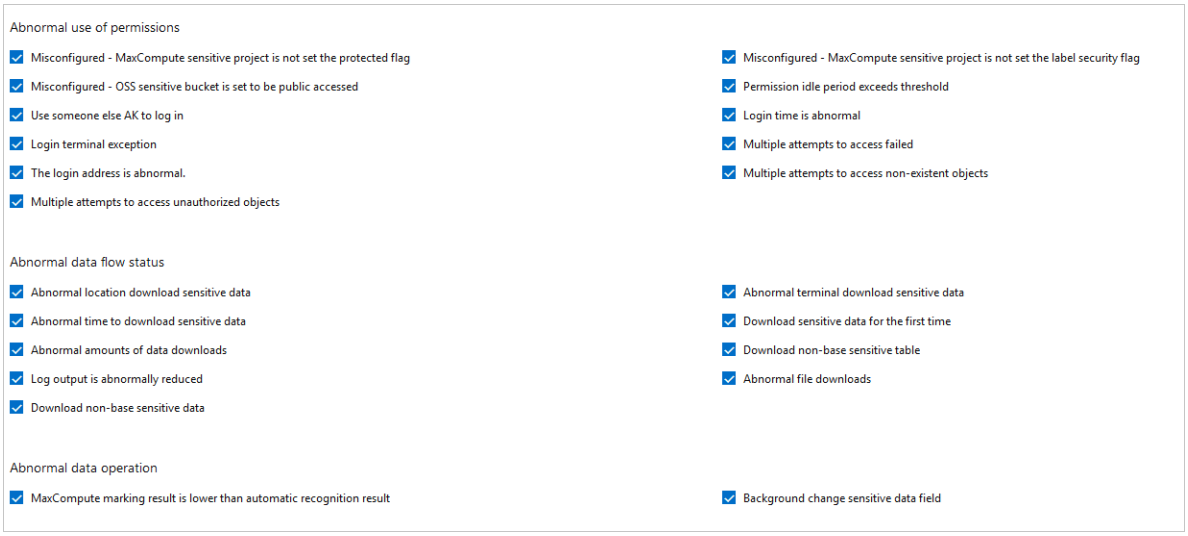
1. *Log on to Apsara Stack Security Center.*
2. **Choose Data Security > Sensitive Data Discovery and Protection > Rule configuration.**
3. **Click the Abnormal output configuration tab to view the thresholds and rules for detecting anomalous activities.**
4. **Configure the thresholds for detecting anomalous activities.**

Sensitive Data Discovery and Protection (SDDP) provides the default threshold values and also allows you to customize the threshold values.

a) **In the Event output threshold configuration section, click Modify next to a threshold.**

b) **Enter a value and click Submit.**

5. Configure the rules for detecting anomalous activities.



In the Time rules applying configuration section, select the types of anomalous activities that you want SDDP to detect. When anomalous activities of the selected types are detected, SDDP displays them on the Abnormal event handling page.

19.8.2.7 Grant access permissions

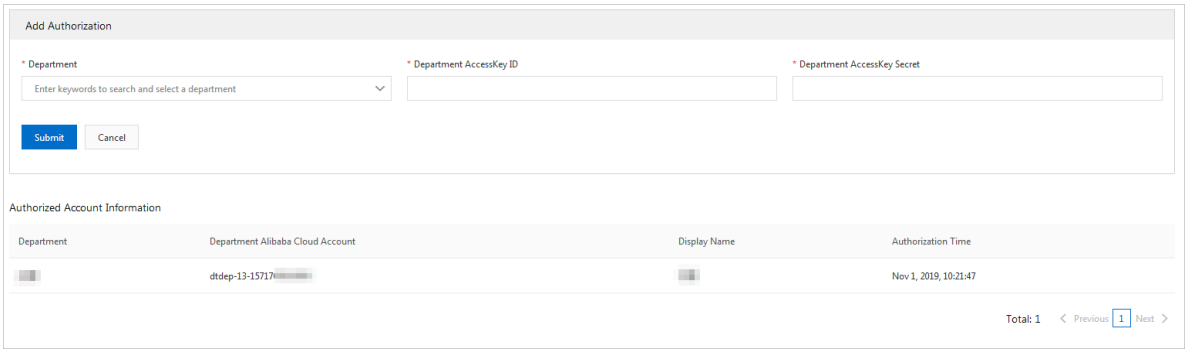
This topic describes how to authorize Sensitive Data Discovery and Protection (SDDP) to access data of your department.

Context

Before using SDDP to detect sensitive data, you must authorize SDDP to access data of Apsara Stack services such as MaxCompute, Object Storage Service (OSS), and Table Store of your department.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Data Security > Sensitive Data Discovery and Protection > Use authorization.



3. In the Add authorization section, authorize SDDP to access data of your department.
   a) In the Department selection drop-down list, enter a keyword and select the department.
   b) Set Department AK ID and Department AK Secret.
   c) Click Submit.
4. In the Authorized account information section, view the list of authorized departments.

20 Key Management Service (KMS)

20.1 What is KMS?

Key Management Service (KMS) is a secure and easy-to-use key management service provided by Apsara Stack. KMS allows you to create and manage CMKs with ease and use a DEKs to encrypt your data.

KMS integrates many Alibaba Cloud products and services to help protect your data in the cloud.

Table 20-1: KMS solutions **describes how KMS provides solutions for a variety of concerns and issues.**

Table 20-1: KMS solutions

Role	Requirement	Solution
Application or website developer	· My program needs keys or certificates for encryption or signature, and I want secure and independent key management services. · I want to securely access keys regardless of where my application is deployed, and cannot take the risk of deploying plaintext keys elsewhere.	KMS provides envelope encryption, allowing you to store the Customer Master Key (CMK) in KMS and deploy only the EDKs. You can simply call a KMS API to decrypt DEKs only when necessary.

Role	Requirement	Solution
Service developer	· I do not want to be responsible for securing users keys and data. · I want users to manage their own keys. I want to use specified keys to encrypt their data after obtaining their authorization. In this way, I can focus on developing service features.	Envelop encryption and KMS APIs allow service developers to use specified CMKs to encrypt and decrypt DEKs. `Plaintexts are not directly stored in a storage device.` This method helps service developers manage CMKs.
Chief security officer (CSO)	· There are compliance requirements that I expect our key management activities to meet. · I need to ensure that keys are reasonably authorized and that the use of any keys is audited.	KMS can connect to RAM for unified authorization management.

20.2 Log on to the KMS console

This topic describes how to log on to the KMS console.

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://*IP address or domain name of the Apsara Stack console*/manage.
- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.

2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click LOGIN to go to the Dashboard page.
5. In the top navigation bar, click the  icon and choose Compute, Storage & Networking > Key Management Service.

20.3 Create a CMK

This topic describes how to create a CMK in the Apsara Stack console for subsequent encryption and decryption operations.

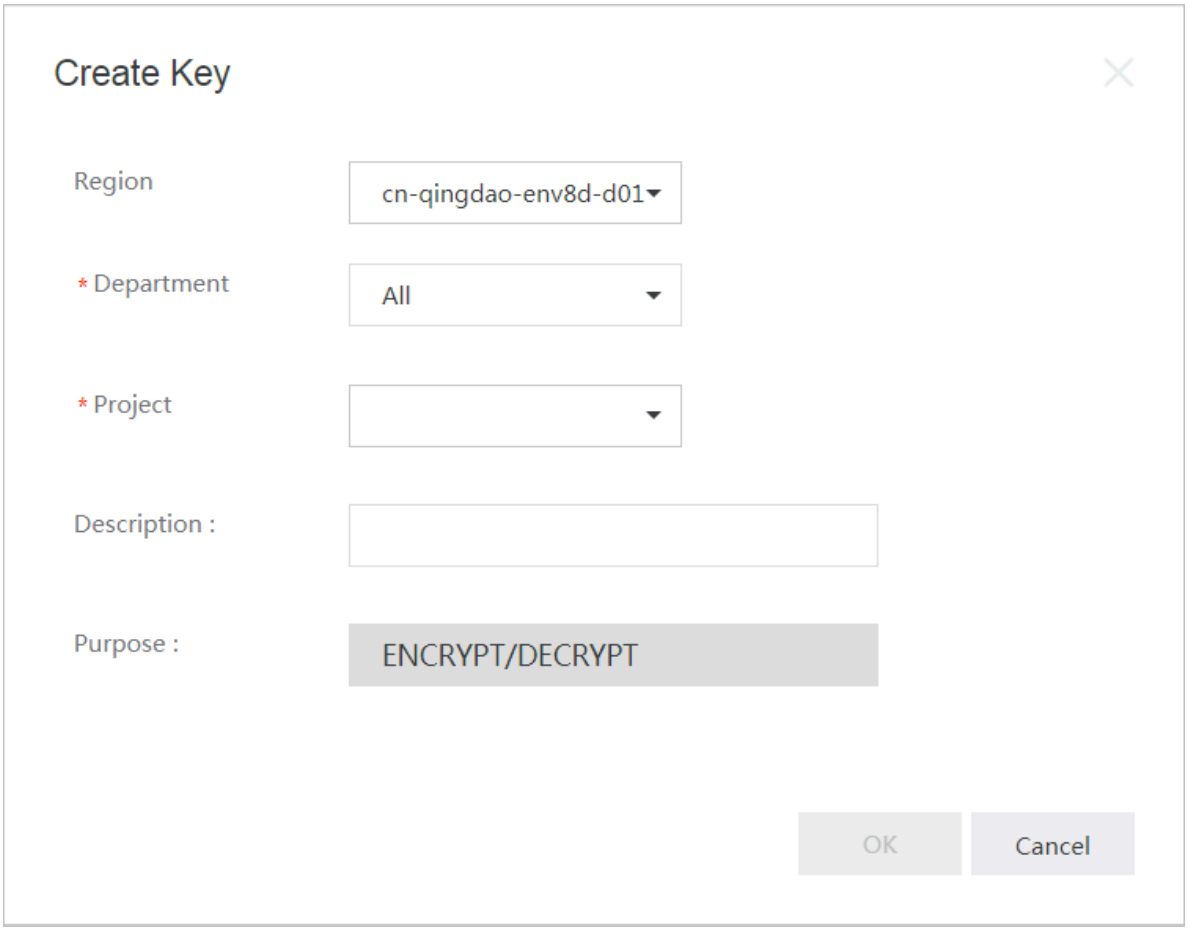
Procedure

1. *Log on to the KMS console*.

2. **Click Create Key.**

   The Create Key dialog box appears, as shown in *Figure 20-1: Create a CMK*.

   Figure 20-1: Create a CMK



3. **Select a region, department, and project. Enter descriptive information. Then, click OK.**

   After the CMK is created, call the KMS API by programming based on Scenarios and the *KMS Developer Guide* **.**

20.4 View CMK details

After a CMK is created, you can view the key ID, key status, key purpose, and creator information.

Procedure

1. *Log on to the KMS console*.

2. In the CMK list, select a CMK that you want to view. Click the link of the key ID, or click the  icon and choose Details from the shortcut menu.

   The Key Details page appears.

3. In Basic Information, you can view the key ID, key status, key purpose, and creator information.

20.5 Enable a CMK

This topic describes how to enable a CMK.

Procedure

1. *Log on to the KMS console*.
2. Select a CMK that is in the `Disabling` state, click the  icon in the Actions column, and choose Enable Key from the shortcut menu.

   After the CMK is enabled, the CMK status changes from `Disabling` to `Enabling`.

20.6 Disable a CMK

If a CMK is disabled, it cannot be used for encryption or decryption. The ciphertext encrypted by using the CMK cannot be decrypted until the CMK is enabled again.

Context

After a CMK is created, it is in the `Enabling` state by default.

Procedure

1. *Log on to the KMS console*.
2. Select a CMK that is in the `Enabling` state, click the  icon, and choose Disable Key from the shortcut menu.

   The Disable Key message appears.

3. Click OK to disable the CMK.

   After the CMK is disabled, the CMK status changes from `Enabling` to `Disabling`.

20.7 Schedule a CMK to be deleted

You can schedule a CMK to be deleted after a specified period from 7 to 30 days.

Context

To delete a CMK, you must specify a scheduled period. The period ranges from 7 to 30 days.

You can set `Cancel Key Deletion` to cancel the CMK deletion before the scheduled period expires.

Note:

- **Within the CMK scheduled deletion period, the CMK is in the Pending Deletion state and cannot be used for operations such as encryption, decryption, and DEK generation.**
- **Deleting a CMK has a severe impact on data availability. Typically, we recommend that you select *Disable a CMK*.**
- **A CMK cannot be recovered after it is deleted. The encrypted content along with the DEK generated by using the CMK cannot be decrypted. Therefore, you can schedule a CMK to be deleted instead of directly deleting it.**
- **KMS deletes the CMK within 24 hours after the scheduled period.**

  For example, if you schedule a CMK to be deleted at 14:00, September 10, 2017, and the scheduled period is seven days, KMS deletes the CMK within 24 hours after 14:00, September 17, 2017.

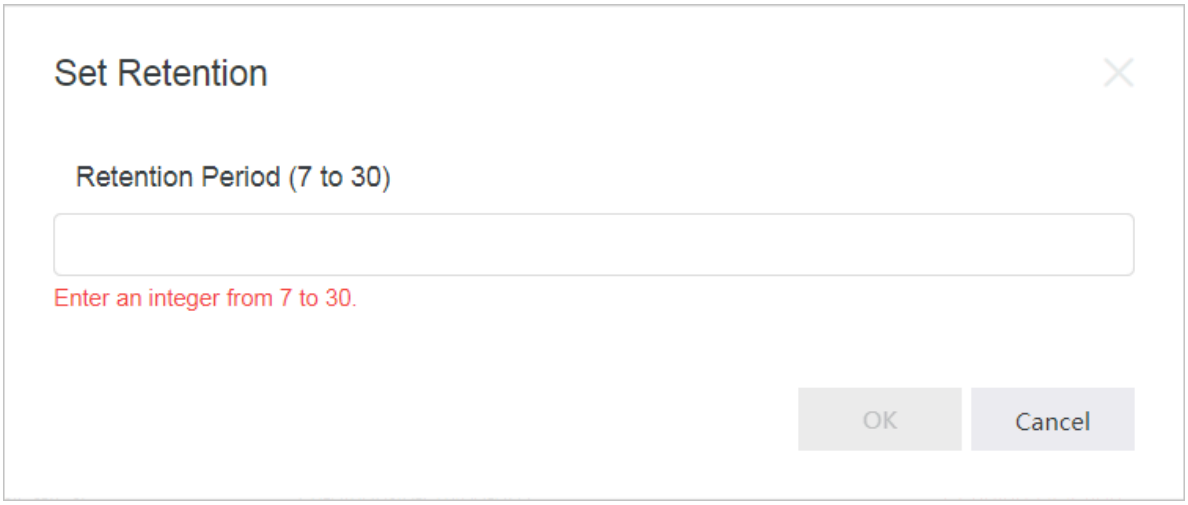
Procedure

1. *Log on to the KMS console*.

2. **Locate a CMK, click the icon in the Actions column, and choose Plan to Delete Key from the shortcut menu.**

   The Plan to Delete Key dialog box appears, as shown in *Figure 20-2: Schedule a CMK to be deleted*.

   Figure 20-2: Schedule a CMK to be deleted



3. **Enter the scheduled period (in days) in the text box, and click OK.**

   Then the CMK status becomes `Pending Deletion`.

   If you want to cancel the scheduled deletion before the scheduled period expires, click the icon in the Actions column and choose Cancel Key Deletion from the shortcut menu.

21 Apsara Stack DNS

21.1 What is Apsara Stack DNS?

Apsara Stack DNS provides basic domain name translation and scheduling services for VPC environments. You can perform the following operations through Apsara Stack DNS in your VPC:

- **Access other ECS servers deployed in VPCs.**
- **Access cloud service instances provided by Apsara Stack.**
- **Access custom enterprise business systems.**
- **Access Internet services and business.**
- **Establish network connections between Apsara Stack DNS and user-created DNS through a leased line.**

You can perform the following operations in the Apsara Stack DNS console:

- **Manage internal domain names**
- **Manage resource record sets of internal domain names**
- **Manage forwarding configurations**
- **Manage recursive resolution**

21.2 User roles and permissions

User role	Permission
System administrator	Owners of this role have read, write, and execute permissions on all level-1 department resources, global resources, and system configurations.
Level-1 department administrator	Owners of this role have read, write, and execute permissions on resources that belong to the level-1 department. However, these owners do not have permissions on resources of other level-1 departments, global resources, and system configurations.

User role	Permission
Lower-level department administrator	Owners of this role do not have permissions on the DNS console or level-1 department resources, global resources, and system configurations.
Resource user	Owners of this role do not have permissions on the DNS console or level-1 department resources, global resources, and system configurations.
Other roles	Owners of this role do not have permissions on the DNS console or level-1 department resources, global resources, and system configurations.

21.3 Log on to the DNS console

This topic uses Google Chrome as an example to demonstrate how to log on to the DNS console.

Prerequisites

- **Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://***IP address or domain name of the Apsara Stack console***/manage.**
- **We recommend that you use the Chrome browser.**

Procedure

1. **Open your browser.**
2. **In the address bar, enter the access address of the Apsara Stack console in the format of http://***IP address or domain name of the Apsara Stack console***/manage, and then press Enter.**
3. **Enter the correct username and password.**
   - **The system has a default super administrator with the username super. The super administrator can create system administrators who can create other system users and notify them of their default passwords by SMS or email.**
   - **You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the**

password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).

4. Click LOGIN to go to the Dashboard page.
5. In the top navigation bar, click and choose Compute, Storage & Networking > DNS.

21.4 Manage internal domain names

21.4.1 Manage tenant internal domain names (Standard Edition only)

21.4.1.1 View tenant internal domain names

Procedure

1. *Log on to the DNS console*.
2. Choose Internal Domains > Tenant Internal Domains.
3. Select a department from the Department drop-down list, or enter a domain name in the Domain Name search box.
4. Click Search.

   The search results are displayed.

21.4.1.2 Create a tenant internal domain name

Procedure

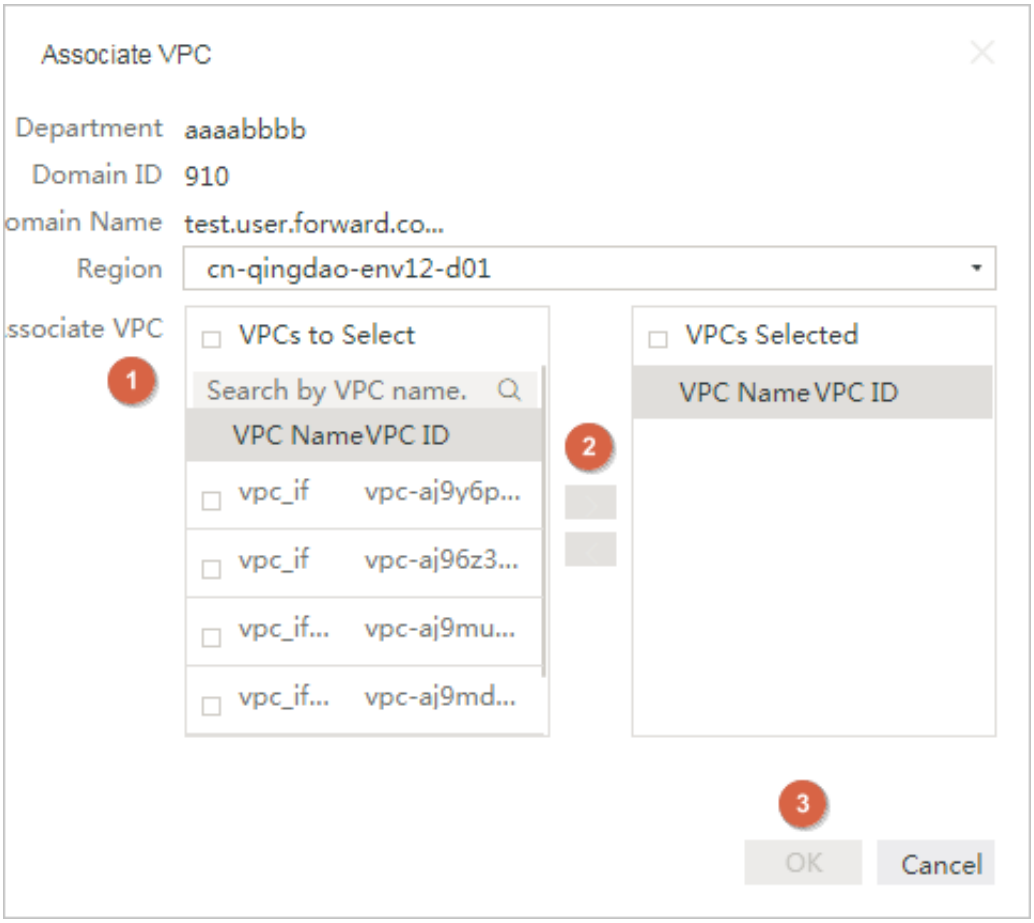
1. *Log on to the DNS console*.
2. Choose Internal Domains > Tenant Internal Domains.
3. Click Create Domain Name.
4. In the dialog box that appears, select a `department`, and enter a `tenant internal domain name`.
5. Click OK.

21.4.1.3 Associate a domain name with a VPC

Tenants are isolated based on VPCs. To ensure that DNS forwarding configurations of a domain name take effect in a VPC, you must associate the domain name with the VPC.

Procedure

1. *Log on to the DNS console*.
2. **Choose Internal Domains > Tenant Internal Domains.**
3. **Find the domain name that you want to associate with a VPC, click the icon in the Actions column, and choose Associate VPC from the shortcut menu.**
4. **Select the VPC that you want to associate from VPCs to Select, click the right arrow button to add the VPC to VPCs Selected, and click OK.**



21.4.1.4 Disassociate a domain name with a VPC

You can disassociate a domain name with a VPC.

Procedure

1. *Log on to the DNS console*.
2. **Choose Internal Domains > Tenant Internal Domains.**

3. **Find the domain name that you want to disassociate with a VPC and click the number in the VPCs Associated column.**
4. **On the VPCs Associated page, find the VPC that you want to disassociate, click the icon in the Actions column, and choose Disassociate VPC from the shortcut menu.**

   The disassociated VPC will not be displayed on the VPCs Associated page.

21.4.1.5 Configure a description for a domain name

Procedure

1. *Log on to the DNS console*.
2. **Choose Internal Domains > Tenant Internal Domains.**
3. **Click the icon in the Actions column of a domain name and choose Description from the shortcut menu.**
4. **In the dialog box that appears, enter a description.**
5. **Click OK.**

21.4.1.6 Delete a domain name

Procedure

1. *Log on to the DNS console*.
2. **Choose Internal Domains > Tenant Internal Domains.**
3. **Click the icon in the Actions column of a domain name and choose Delete from the shortcut menu.**
4. **In the message that appears, click OK.**

21.4.1.7 Delete multiple domain names

Procedure

1. *Log on to the DNS console*.
2. **Choose Internal Domains > Tenant Internal Domains.**
3. **Select one or more domain names that you want to delete, and click Delete Domain Names in the upper-right corner.**
4. **In the message that appears, click OK.**

21.4.1.8 Manage resource records

Procedure

1. *Log on to the DNS console*.
2. Choose Internal Domains > Tenant Internal Domains.
3. Find the domain name for which you want to manage resource records, click the icon in the Actions column, and choose Manage Resource Records from the shortcut menu.
4. On the Manage Resource Records page, click Add Resource Record Set in the upper-right corner.
5. In the Add Resource Record Set dialog box that appears, specify a `hostname`, `record type`, `resolution policy`, and `TTL`. Enter a resource record set in the `Data` field, and click OK.

   The formatting rules of different record types are described as follows:

   - A record

Resolution policy	Formatting rule
None	Enter each IPv4 address in a single row. You can enter up to 100 IPv4 addresses in separate rows. You cannot enter repeated IPv4 addresses. IPv4 addresses must be written in standard IPv4 address format. Example: - 192.168.1.1 - 192.168.1.2 - 192.168.1.3

Resolution policy	Formatting rule
Weight	Enter each IPv4 address in a single row. You can enter up to 100 IPv4 addresses in separate rows. You cannot enter repeated IPv4 addresses. Format: - [IPv4 address] [Weight]. Separate the IPv4 address and weight with a space. - IPv4 addresses must be written in standard IPv4 address format. - The value of a weight must be an integer in the range of 0 to 999. A greater value indicates a heavier weight. Example: - 192.168.1.1 20 - 192.168.1.1 30 - 192.168.1.1 50

- AAAA record

Resolution policy	Formatting rule
None	Enter each IPv6 address in a single row. You can enter up to 100 IPv6 addresses in separate rows. You cannot enter repeated IPv6 addresses. IPv6 addresses must be written in standard IPv6 address format. Example: - 2400:3200::6666 - 2400:3200::6688 - 2400:3200::8888

Resolution policy	Formatting rule
Weight	Enter each IPv6 address in a single row. You can enter up to 100 IPv6 addresses in separate rows. You cannot enter repeated IPv6 addresses. Format: - [IPv6 address] [Weight]. Separate the IPv6 address and weight with a space. - IPv6 addresses must be written in standard IPv6 address format. - The value of a weight must be an integer in the range of 0 to 999. A greater value indicates a heavier weight. Example: - 2400:3200::6666 20 - 2400:3200::6688 20 - 2400:3200::8888 60

- CNAME record

Resolution policy	Formatting rule
None	You can enter only one domain name, and enter the domain name in a single row. The domain name must be a fully qualified domain name (FQDN). It must end with a period (.). It must be 1 to 255 ASCII characters in length. Example: www.example.com.

Resolution policy	Formatting rule
Weight	Enter each domain name in a single row. You can enter up to 100 domain names in separate rows. You cannot enter repeated domain names. Format: - [Domain name] [Weight]. Separate the domain name and weight with a space. - The domain name must be a fully qualified domain name (FQDN). It must end with a period (.). It must be 1 to 255 ASCII characters in length. - The value of a weight must be an integer in the range of 0 to 999. A greater value indicates a heavier weight. Example: - www1.example.com. 20 - www2.example.com. 20 - www3.example.com. 60

- MX record

Resolution policy	Formatting rule
None	Enter each MX record in a single row. You can enter up to 100 MX records in separate rows. You cannot enter repeated MX records. Format: - [Priority] [Email server hostname]. Separate the priority and email server hostname with a space. - The value of a priority must be an integer in the range of 0 to 999. A smaller value indicates a higher priority. - The email server hostname must be a fully qualified domain name (FQDN). It must end with a period (.). It must be 1 to 255 ASCII characters in length. Example: - 10 mailserver1.example.com. - 20 mailserver2.example.com

- TXT record

Resolution policy	Formatting rule
None	Enter each TXT record in a single row. You can enter up to 100 TXT records in separate rows. You cannot enter repeated TXT records. A TXT record must be 1 to 255 ASCII characters in length . Each row cannot be blank. Example: "v=spf1 ip4:192.168.0.1/16 ip6:2001::1/96 ~all"

- PTR record

Resolution policy	Formatting rule
None	Enter each domain name in a single row. You can enter up to 100 domain names in separate rows. You cannot enter repeated domain names. The domain name must be a fully qualified domain name (FQDN). It must end with a period (.). It must be 1 to 255 ASCII characters in length. Example: - www1.example.com. - www2.example.com. - www3.example.com.

- SRV record

Resolution policy	Formatting rule
None	Enter each SRV record in a single row. You can enter up to 100 SRV records in separate rows. You cannot enter repeated SRV records. Format: - [Priority] [Weight] [Port number] [Application server hostname]. Separate the priority, weight, port number , and application server hostname with spaces. - The value of a priority must be an integer in the range of 0 to 999. A smaller value indicates a higher priority. - The value of a weight must be an integer in the range of 0 to 999. A greater value indicates a heavier weight. - A port number must be an integer in the range of 0 to 65535. It indicates the TCP or UDP port that is used for network communication. - The application server hostname must be a fully qualified domain name (FQDN). It must end with a period (.). It must be 1 to 255 ASCII characters in length. Example: - 1 10 8080 www1.example.com. - 2 20 8081 www2.example.com.

- NAPTR record

Resolution policy	Formatting rule
None	Enter each NAPTR record in a single row. You can enter up to 100 NAPTR records in separate rows. You cannot enter repeated NAPTR records. Format: - [Serial number] [Priority] [Flag] [Service information] [Regular expression] [Substitute domain name]. Separate the serial number, priority, flag, service information, regular expression, and substitute domain name with spaces. - A serial number must be an integer in the range of 0 and 999. A smaller serial number indicates a higher priority. - The value of a priority must be an integer in the range of 0 to 999. When multiple records contain the same serial number, records that have a higher priority take precedence. - The value of a flag can be null or a single character in the range of A to Z, a to z, or 0 to 9. It is not case-sensitive and must be enclosed in double quotation marks ("). - The value of the service information can be null and must be 1 to 32 ASCII characters in length. It must start with a letter and be enclosed in double quotation marks ("). - The value of the regular expression can be null and must be 1 to 255 ASCII characters in length. It must be enclosed in double quotation marks ("). - The substitute domain name must be a fully qualified domain name (FQDN). It must end with a period (.). It must be 1 to 255 ASCII characters in length. Example: - 100 50 "S" "Z3950+I2L+I2C" "" _z3950._tcp.example.com.
	- 100 50 "S" "RCDS+I2C" "" _rcds._udp.example.com. - 100 50 "S" "HTTP+I2L+I2C+I2R" "" _http._tcp.example.com.

- **CAA record**

Resolution policy	Formatting rule
None	Enter each CAA record in a single row. You can enter up to 100 CAA records in separate rows. You cannot enter repeated CAA records. Format: - [Certification authority flag] [Certificate property tag] [Authorization information]. Separate the certification authority flag, certificate property tag, and authorizat ion information with spaces. - The certification authority flag must be an integer. It must be 0 to 255 characters in length. - The value of the certificate property tag can be issue, issuewild, and iodef. - The value of the authorization information cannot be null and must be 1 to 255 ASCII characters in length. It must be enclosed in double quotation marks ("). Example: - 0 issue "caa.example.com" - 0 issuewild ";" - 0 iodef "mailto:example@example.com"

- **NS record**

Resolution policy	Formatting rule
None	Enter each DNS server address in a single row. You can enter up to 100 DNS server addresses in separate rows. You cannot enter repeated DNS server addresses. The DNS server address must be a fully qualified domain name (FQDN). It must end with a period (.). It must be 1 to 255 ASCII characters in length. Wildcard subdomains are not allowed. Example: - ns1.example.com. - ns2.example.com.

6. After you have added resource record sets, you can perform the following operations:
   - Configure a description for a resource record set

     Find the resource record set for which you want to configure a description, click the icon in the Actions column, and choose Description from the shortcut menu. Enter a description and click OK.
   - Delete a resource record set

     Find the resource record set that you want to delete, click the icon in the Actions column, and choose Delete from the shortcut menu. In the message that appears, click OK.
   - Change a resource record set

     Click the icon in the Actions column of a resource record set and choose Change from the shortcut menu. In the dialog box that appears, enter the required information and click OK.
   - Delete multiple resource record sets

     Select the resource record sets that you want to delete, and click Delete in the upper-right corner. In the message that appears, click OK.

21.4.1.9 View the resolution policy

You can view the details of the resolution policy.

Procedure

1. *Log on to the DNS console*.
2. Choose Internal Domains > Tenant Internal Domains.
3. Click the icon in the Actions column of a domain name, and choose Manage Resource Records from the shortcut menu.
4. Click Weight in the Resolution Policy column of a resource record set. The weight and percentage of each resource record are displayed.

21.4.2 Global internal domain names

21.4.2.1 Overview

All operations of this feature require administrative privileges.

21.4.2.2 View global internal domain names

Procedure

1. *Log on to the DNS console*.
2. Choose Internal Domains > Global Internal Domains.
3. Enter a domain name in the Domain Name search box.
4. Click Search.

   The search results are displayed.

21.4.2.3 Create a global internal domain name

You can create a global internal domain name in the Apsara Stack console.

Procedure

1. *Log on to the DNS console*.
2. Choose Internal Domains > Global Internal Domains.
3. Click Create Domain Name.
4. In the dialog box that appears, enter a global internal domain name.
5. Click OK.

21.4.2.4 Configure a description for a domain name

You can configure a description for a domain name in the Apsara Stack console.

Context

An informative description such as the hostname and internal information system helps you understand the purpose of the domain name.

Procedure

1. *Log on to the DNS console*.
2. **Choose Internal Domains > Global Internal Domains.**
3. **Click the icon in the Actions column of a domain name and choose Description from the shortcut menu.**
4. **In the dialog box that appears, enter a description.**
5. **Click OK.**

21.4.2.5 Delete a domain name

You can delete a domain name in the Apsara Stack console.

Procedure

1. *Log on to the DNS console*.
2. **Choose Internal Domains > Global Internal Domains.**
3. **Click the icon in the Actions column of a domain name and choose Delete from the shortcut menu.**
4. **In the message that appears, click OK.**

21.4.2.6 Delete multiple domain names

You can delete multiple domain names in the Apsara Stack console.

Procedure

1. *Log on to the DNS console*.
2. **Choose Internal Domains > Global Internal Domains.**
3. **Select the domain names that you want to delete and click Delete Domain Names in the upper-right corner.**
4. **In the message that appears, click OK.**

21.4.2.7 Manage resource records

You can manage resource records in the Apsara Stack console.

Procedure

1. *Log on to the DNS console*.
2. **Choose Internal Domains > Global Internal Domains.**

3. Find the domain name for which you want to manage resource records, click the icon in the Actions column, and choose Manage Resource Records from the shortcut menu.
4. On the Manage Resource Records page, click Add Resource Record Set in the upper-right corner.
5. After you have added resource record sets, you can perform the following operations:
   - Configure a description for a resource record set

     Find the resource record set for which you want to configure a description, click the management icon in the Actions column, and choose Description from the shortcut menu. Enter a description and click OK.
   - Delete a resource record set

     Find the resource record set that you want to delete, click the management icon in the Actions column, and choose Delete from the shortcut menu. In the message that appears, click OK.
   - Change a resource record set

     Click the management icon in the Actions column of a resource record set and choose Change from the shortcut menu. In the dialog box that appears, enter the required information and click OK.
   - Delete multiple resource record sets

     Select the resource record sets that you want to delete, and click Delete in the upper-right corner. In the message that appears, click OK.

21.4.2.8 View the resolution policy

Procedure

1. *Log on to the DNS console*.
2. Choose Internal Domains > Global Internal Domains.
3. Click the icon in the Actions column of a domain name and choose Manage Resource Records from the shortcut menu.
4. Click Weight in the Resolution Policy column of a resource record set.
5. On the Resolution Policy page, the weight and percentage of each resource record are displayed.

21.5 Forwarding configurations

21.5.1 Tenant forwarding configurations (Standard Edition only)

21.5.1.1 Tenant domain names

21.5.1.1.1 View tenant domain names

Procedure

1. *Log on to the DNS console*.
2. **Choose Forwarding Configurations > Tenant Forwarding Configurations > Tenant Forward Domains.**
3. **Select a department from the Department drop-down list, or enter a domain name in the Domain Name search box.**
4. **Click Search.**

   The search results are displayed.

21.5.1.1.2 Add a domain name

Procedure

1. *Log on to the DNS console*.
2. **Choose Forwarding Configurations > Tenant Forwarding Configurations > Tenant Forward Domains.**
3. **Click Create Domain Name.**
4. **In the dialog box that appears, select a** `department`**, enter a** `domain name`**, select a** `forwarding mode`**, and enter one or more** `forwarder IP addresses`**.**

Name	Description
Department	The department to which the domain name belongs.

Name	Description
Domain Name	The domain name. The formatting rules of domain names are described as follows: · A domain name must be 1 to 255 characters in length, including the period (.) at the end of the domain name. · A domain name can contain multiple domain name segments that are separated with periods (.). A domain name segment must be 1 to 63 characters in length and cannot be empty. It cannot contain consecutive periods (.). · A domain name can contain letters (a–z, A–Z), numbers (0–9), hyphens (-), and underscores (_). Domain names that contain other characters are invalid. · A domain name must start with a letter, number, or underscore (_) and must end with a letter, number, or period (.). · Domain names are not case-sensitive. The system saves domain names in lowercase letters by default. · The period (.) at the end of a domain name is optional. The system adds a period (.) to the end of a domain name that does not end with a period (.).
Forwarding Mode	For both domain name-based forwarding and default forwarding, two forwarding modes are available: forward all requests with recursion and forward all requests without recursion. · Forward all requests without recursion: forwards DNS requests to the target DNS server. If the target DNS server cannot resolve the domain names, a message is returned to the DNS client indicating that the query failed. · Forward all requests with recursion: forwards DNS requests to the target DNS server. If the target DNS server cannot resolve the domain names, the local DNS server is used to resolve them. Note: If you enter private IP addresses in the Forwarder IP Addresses field, unexpected results may occur during recursive resolution. For example, a domain name used for private network services may be resolved to a public IP address.

Name	Description
Forwarder IP Addresses	The IP addresses of target DNS servers. **Note:** Separate multiple IP addresses with semicolons (;).

5. Click OK.

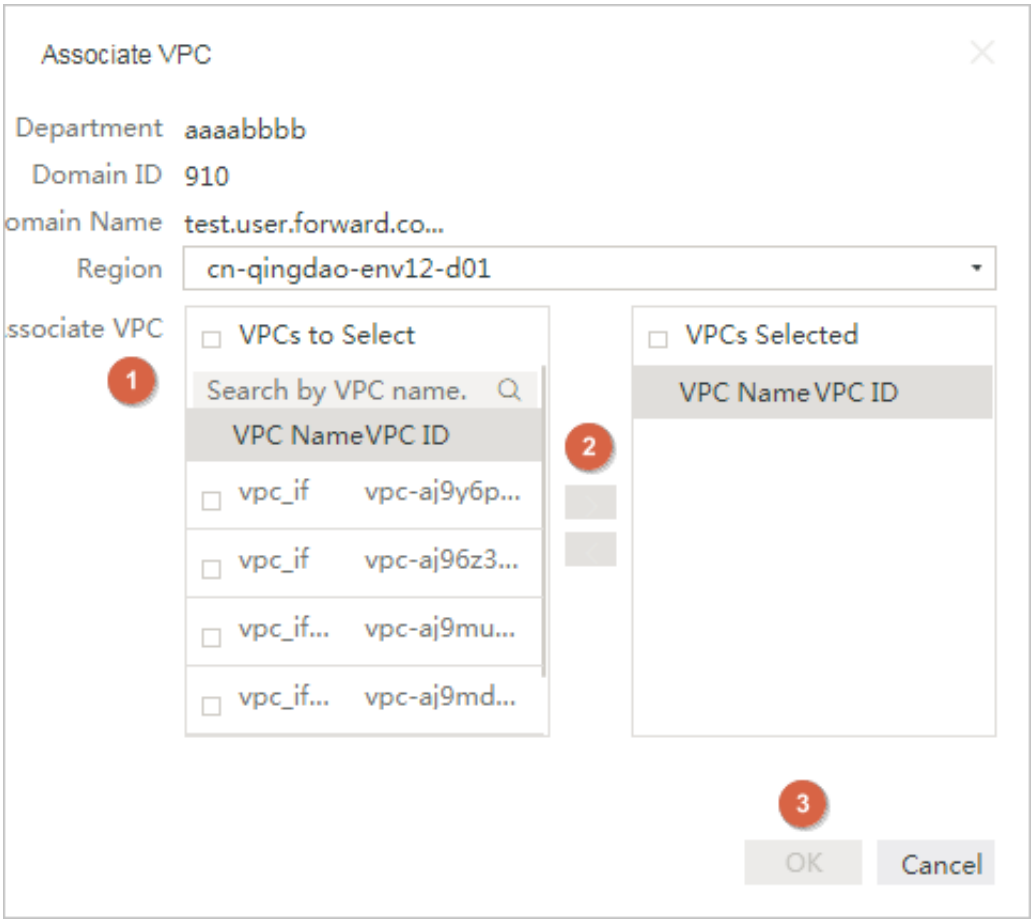
21.5.1.1.3 Associate a domain name with a VPC

Tenants are isolated based on VPCs. To ensure that DNS forwarding configurations of a domain name take effect in a VPC, you must associate the domain name with the VPC.

Procedure

1. *Log on to the DNS console*.
2. Choose Forwarding Configurations > Tenant Forwarding Configurations > Tenant Forward Domains.
3. Find the domain name that you want to associate with a VPC, click the icon in the Actions column, and choose Associate VPC from the shortcut menu.

4. **Select the VPC that you want to associate from VPCs to Select, click the right arrow button to add the VPC to VPCs Selected, and click OK.**



21.5.1.1.4 Disassociate a domain name with a VPC

You can disassociate a domain name with a VPC.

Procedure

1. *Log on to the DNS console*.
2. **Choose Forwarding Configurations > Tenant Forwarding Configurations > Tenant Forward Domains.**
3. **Find the domain name that you want to disassociate with a VPC and click the number in the VPCs Associated column.**
4. **On the VPCs Associated page, find the VPC that you want to disassociate, click the icon in the Actions column, and choose Disassociate VPC from the shortcut menu.**

   The disassociated VPC will not be displayed on the VPCs Associated page.

21.5.1.1.5 Change forwarding configurations of a domain name

Procedure

1. *Log on to the DNS console*.
2. **Choose Forwarding Configurations > Tenant Forwarding Configurations > Tenant Forward Domains.**
3. **Click the icon in the Actions column of a domain name and choose Change from the shortcut menu.**
4. **In the dialog box that appears, modify the forwarding mode or forwarder IP addresses.**
5. **Click OK.**

21.5.1.1.6 Configure a description for a domain name

Procedure

1. *Log on to the DNS console*.
2. **Choose Forwarding Configurations > Tenant Forwarding Configurations > Tenant Forward Domains.**
3. **Click the icon in the Actions column of a domain name and choose Description from the shortcut menu.**
4. **In the dialog box that appears, enter a description.**
5. **Click OK.**

21.5.1.1.7 Delete a domain name

Procedure

1. *Log on to the DNS console*.
2. **Choose Forwarding Configurations > Tenant Forwarding Configurations > Tenant Forward Domains.**
3. **Click the icon in the Actions column of a domain name and choose Delete from the shortcut menu.**
4. **In the message that appears, click OK.**

21.5.1.1.8 Delete domain names

Procedure

1. *Log on to the DNS console*.
2. Choose Forwarding Configurations > Tenant Forwarding Configurations > Tenant Forward Domains.
3. Select one or more domain names that you want to delete, and click Delete Domain Names in the upper right corner.
4. In the message that appears, click OK.

21.5.1.2 Default forwarding configurations

21.5.1.2.1 View default forwarding configurations

Prerequisites

Only system administrators and level-1 department administrators are authorized to perform this operation.

Procedure

1. *Log on to the DNS console*.
2. Choose Forwarding Configurations > Tenant Forwarding Configurations > Default Forwarding Configurations.
3. Select your department from the Department drop-down list.
4. Click Search.

   The search results are displayed.

21.5.1.2.2 Add a default forwarding configuration

Prerequisites

Only the system administrators and the level-1 department administrators are authorized to perform this operation.

Procedure

1. *Log on to the DNS console*.
2. Choose Forwarding Configurations > Tenant Forwarding Configurations > Default Forwarding Configurations.
3. Click Add Configuration in the upper right corner.

4. In the dialog box that appears, select a `department`, select a `forwarding mode`, and enter one or more IP addresses in the `Forwarder IP Addresses` field.

Name	Description
Department	The department to which the default forwarding configuration belongs.
Forwarding Mode	For both domain name-based forwarding and default forwarding, two forwarding modes are available: forward all requests with recursion and forward all requests without recursion. · Forward all requests without recursion: forwards DNS requests to the target DNS server. If the target DNS server cannot resolve the domain names, a message is returned to the DNS client indicating that the query failed. · Forward all requests with recursion: forwards DNS requests to the target DNS server. If the target DNS server cannot resolve the domain names, the local DNS server is used to resolve them. Note: If you enter private IP addresses in the Forwarder IP Addresses field, unexpected results may occur during recursive resolution. For example, a domain name used for internal network services may be resolved to a public IP address.
Forwarder IP Addresses	IP addresses of target DNS servers. Note: Multiple IP addresses are separated with semicolons (;).

5. Click OK.

21.5.1.2.3 Associate a forwarding configuration with a VPC

Tenants are isolated based on VPCs. To ensure that a DNS forwarding configuration takes effect in a VPC, you must associate the forwarding configuration with the VPC.

Prerequisites

Only system administrators and level-1 department administrators are authorized to perform this operation.

Procedure

1. *Log on to the DNS console*.
2. **Choose Forwarding Configurations > Tenant Forwarding Configurations > Default Forwarding Configurations.**
3. **Find the configuration that you want to associate with a VPC, click the icon in the Actions column, and choose Associate VPC from the shortcut menu.**
4. **Select the VPC that you want to associate from VPCs to Select, click the right arrow button to add the VPC to VPCs Selected, and click OK.**

21.5.1.2.4 Disassociate a domain name with a VPC

You can disassociate a domain name with a VPC.

Procedure

1. *Log on to the DNS console*.
2. **Choose Forwarding Configurations > Tenant Forwarding Configurations > Default Forwarding Configurations.**
3. **Find the domain name that you want to disassociate with a VPC and click the number in the VPCs Associated column.**
4. **On the VPCs Associated page, find the VPC that you want to disassociate, click the icon in the Actions column, and choose Disassociate VPC from the shortcut menu.**

   The disassociated VPC will not be displayed on the VPCs Associated page.

21.5.1.2.5 Modify a default forwarding configuration

Prerequisites

Only system administrators and level-1 department administrators are authorized to perform this operation.

Procedure

1. *Log on to the DNS console*.
2. **Choose Forwarding Configurations > Tenant Forwarding Configurations > Default Forwarding Configurations.**

3. Click the icon in the Actions column of a forwarding configuration and choose Change from the shortcut menu.
4. In the dialog box that appears, modify the forwarding mode or forwarder IP addresses.
5. Click OK.

21.5.1.2.6 Configure a description for a default forwarding configuration

Prerequisites

Only system administrators and level-1 department administrators are authorized to perform this operation.

Procedure

1. *Log on to the DNS console*.
2. Choose Forwarding Configurations > Tenant Forwarding Configurations > Default Forwarding Configurations.
3. Click the icon in the Actions column of a forwarding configuration and choose Description from the shortcut menu.
4. In the dialog box that appears, enter a description.
5. Click OK.

21.5.1.2.7 Delete a default forwarding configuration

Prerequisites

Only system administrators and level-1 department administrators are authorized to perform this operation.

Procedure

1. *Log on to the DNS console*.
2. Choose Forwarding Configurations > Tenant Forwarding Configurations > Default Forwarding Configurations.
3. Click the icon in the Actions column of a forwarding configuration and choose Delete from the shortcut menu.
4. In the message that appears, click OK.

21.5.1.2.8 Delete default forwarding configurations

Prerequisites

Only system administrators and level-1 department administrators are authorized to perform this operation.

Procedure

1. *Log on to the DNS console*.
2. **Choose Forwarding Configurations > Tenant Forwarding Configurations > Default Forwarding Configurations.**
3. **Select one or more forwarding configurations that you want to delete and click Delete Forwarding Configurations in the upper-right corner.**
4. **In the message that appears, click OK.**

21.5.2 Global forwarding configuration

21.5.2.1 Global forwarding domains

21.5.2.1.1 Overview

All operations of this feature require administrative privileges.

Apsara Stack DNS can forward DNS requests for specific domain names to other DNS servers.

Two forwarding modes are available: forward all requests with recursion and forward all requests without recursion.

- **Forward all requests without recursion: forwards DNS requests to the target DNS server. If the target DNS server cannot resolve the domain names or the request is timed out, a message is returned to the DNS client indicating that the query failed.**
- **Forward all requests with recursion: forwards DNS requests to the target DNS server. If the target DNS server cannot resolve the domain names, the local DNS server is used to resolve them.**

21.5.2.1.2 View forwarding rules of a domain name

You can search for a domain name and view its forwarding rules in the Apsara Stack console. This operation requires administrative rights.

Procedure

1. *Log on to the DNS console*.

2. **Choose Forwarding Configurations > Global Forwarding Configuration > Global Forward Domains.**
3. **Enter a domain name in the Domain Name search box and click Search to view forwarding configurations of the domain name.**

21.5.2.1.3 Create a domain name

You can create a domain name in the Apsara Stack console. This operation requires administrative rights.

Procedure

1. *Log on to the DNS console*.
2. **Choose Forwarding Configurations > Global Forwarding Configuration > Global Forward Domains.**
3. **Click Create Domain Name.**
4. **In the dialog box that appears, enter a** `domain name`**, select a** `forwarding mode`**, enter one or more** `forwarder IP addresses`**, and click OK.**

21.5.2.1.4 Configure a description for a domain name

You can configure a description for a domain name in the Apsara Stack console. This operation requires administrative rights.

Context

A description such as the hostname and internal information system helps you understand the purpose of the domain name.

Procedure

1. *Log on to the DNS console*.
2. **Choose Forwarding Configurations > Global Forwarding Configuration > Global Forward Domains.**
3. **Find the domain name for which you want to configure a description, click the**  **icon in the Actions column, and choose Description from the shortcut menu.**
4. **In the dialog box that appears, enter a description, and click OK.**

21.5.2.1.5 Change forwarding configurations for a domain name

You can change forwarding configurations for a domain name in the Apsara Stack console. This operation requires administrative rights.

Procedure

1. *Log on to the DNS console*.
2. **Choose Forwarding Configurations > Global Forwarding Configuration > Global Forward Domains.**
3. **Find the domain name for which you want to change forwarding configurations, click the icon in the Actions column, and choose Change from the shortcut menu.**
4. **In the dialog box that appears, select a** `forwarding mode`**, enter one or more** `forwarder IP addresses`**, and click OK.**

21.5.2.1.6 Delete a domain name

You can delete a domain name in the Apsara Stack console. This operation requires administrative rights.

Procedure

1. *Log on to the DNS console*.
2. **Choose Forwarding Configurations > Global Forwarding Configuration > Global Forward Domains.**
3. **Select the domain name that you want to delete, click the icon in the Actions column, and choose Delete from the shortcut menu.**
4. **Click OK.**

21.5.2.1.7 Delete domain names

You can delete one or more domain names in the Apsara Stack console. This operation requires administrative rights.

Procedure

1. *Log on to the DNS console*.
2. **Choose Forwarding Configurations > Global Forwarding Configuration > Global Forward Domains.**
3. **Select one or more domain names that you want to delete and click Delete Domain Names in the upper-right corner.**

4. Click OK.

21.5.2.2 Global default forwarding configuration

21.5.2.2.1 Enable default forwarding

You can enable default forwarding in the Apsara Stack console. This operation requires administrative rights.

Procedure

1. *Log on to the DNS console*.
2. Choose Forwarding Configurations > Global Forwarding Configuration > Global Default Forwarding Configuration.
3. Click the icon in the Actions column and choose Enable from the shortcut menu.
4. In the dialog box that appears, select a `forwarding mode`, enter one or more `forwarder IP addresses`, and click OK.

   After you enable default forwarding, `Enable Default Forwarding` is set to `ON`.

21.5.2.2.2 Change the global configurations of default forwarding

You can change the global configurations of default forwarding in the Apsara Stack console. This operation requires administrative rights.

Procedure

1. *Log on to the DNS console*.
2. Choose Forwarding Configurations > Global Forwarding Configuration > Global Default Forwarding Configuration.
3. Click the icon in the Actions column and choose Change from the shortcut menu.
4. In the dialog box that appears, select a `forwarding mode`, enter one or more `forwarder IP addresses`, and click OK.

21.5.2.2.3 Disable default forwarding

You can disable default forwarding in the Apsara Stack console. This operation requires administrative rights.

Procedure

1. *Log on to the DNS console*.
2. **Choose Forwarding Configurations > Global Forwarding Configuration > Global Default Forwarding Configuration.**
3. **Click the icon in the Actions column and choose Disable from the shortcut menu.**
4. **Click OK.**

21.6 Recursive resolution

21.6.1 Enable global recursive resolution

Prerequisites

Only the system administrators are authorized to manage recursive resolution.

Procedure

1. *Log on to the DNS console*.
2. **Click the Recursive Resolution tab.**
3. **Click in the Actions column and choose Enable from the shortcut menu.**
4. **Click OK.**

21.6.2 Disable global recursive resolution

Prerequisites

Only the system administrators are authorized to manage recursive resolution.

Procedure

1. *Log on to the DNS console*.
2. **Click the Recursive Resolution tab.**
3. **Click in the Actions column and choose Disable from the shortcut menu.**
4. **Click OK.**